

Los Alamos deserves better

The US Congress has turned the Los Alamos National Laboratory into a political pawn. The damage done will take a long time to rectify. Despite their mistakes, the national defence laboratories need far greater support from their country's leaders.

So the flogging continues until morale improves. That is how a US government scientist characterized the position in which employees of the Los Alamos National Laboratory find themselves. It is a description that is all too apt. Whipping will probably continue for as long as it suits the political agendas of those, in Congress above all, doing the whipping, and at least until the presidential election in November. But the lab's lacerations will cut deep and take far longer to heal. What talented scientist will want to come and work at the laboratory with this lengthy public excoriation echoing in their memory?

Of course Los Alamos deserves criticism. Temporarily missing disks containing sensitive information about nuclear weapons should not have remained unaccounted for as long as they did. But it's worth putting that episode into perspective. The disks, containing information on US and other nuclear weapons, did not have the highest security classification. The people who used them were volunteers for an important job: tackling emergencies involving unexploded nuclear devices. The most plausible scenario for what happened is that one or more such volunteers acted irresponsibly, did not record taking the disks, then found themselves in the middle of a hunt. At some point, they decided to return them by dropping them behind a photocopier. The fact that much of this took place during a major threat to the lab's existence from forest fires, when the staff responded as best they could as homes were destroyed, is surely a mitigating fact to be considered.

The political agendas now in play need to be remembered. The energy secretary Bill Richardson, a partisan figure who, until this episode, was considered a possible running mate for Al Gore, is the real target of the latest bout of hysteria over lab security. Los Alamos and the Lawrence Livermore National Laboratory are caught up in the attacks on him.

But the politics are occurring against a background of longer-term decline in morale at those labs. Since the case of the alleged Chinese spy

Wen Ho Lee arose in 1998, they have faced a growing catalogue of obstacles to getting on with their main work — research needed to sustain the nuclear stockpile in the absence of nuclear testing. Security issues and congressional constraints have led to travel restrictions, scientifically unproven and insulting polygraph tests, obstacles to the hiring of foreign nationals, and cuts in spending at the discretion of lab directors. The growth in bureaucracy and uncertainty in programme status have undermined the labs' attractiveness as places with which to collaborate. Above all, they bear a growing burden of micro-management from congressional committees and their staff.

Of course, the laboratories themselves have not adapted quickly enough to the post-cold-war environment. Huge cost overruns at Lawrence Livermore's national ignition facility for fusion research are an example of how the labs' lack of accountability has undermined their credibility. But the damage now being done to their reputation is disproportionate to their sins. These laboratories are not only essential for the nuclear and, increasingly, biological security of the United States. They can also bring interdisciplinary skills and scales of effort to bear on other important research topics in a way that few other laboratories can.

Earlier this week, Richardson announced that the University of California will be asked to renegotiate its contract for running the labs and thereby lose its responsibility for lab security. But the difficulty of balancing science and security at the labs has been apparent since the start of the Manhattan Project during the Second World War, and many participants in the current debate seem ignorant of the inherent difficulties of running a nuclear weapons research programme in a country as committed to freedom of information as is the United States. The incoming administration will have no option but to shake off the political invective and move rapidly to rebuild public confidence in what remain jewels in the US research crown, whose very future as top-quality research laboratories is now in danger. ■

Declaration for AIDS sufferers

A declaration about HIV as the cause of AIDS should be accepted and thus bring an end to a tragic debate in South Africa.

The "Durban Declaration", which is published as a Commentary on page 15, has been signed by scientists, doctors and senior representatives of all the world's leading medical and scientific research organizations and government authorities engaged in HIV/AIDS research or policy — UNAIDS, medical research institutes, leading universities and independent research foundations. But, crucially, the declaration is also supported by health ministers, AIDS experts and public-health officials from countries in Africa, Asia and elsewhere in the developing world (see page 3).

The declaration is a massive international response to recent debates in South Africa, and is made on behalf of people infected with HIV. Frustrated at a needless and tragic delay in treating sufferers, the authors see the declaration as a decisive rejection of arguments put forward by many of those on a panel set up by South African President Thabo Mbeki to "re-evaluate" the cause of AIDS. As described in the

declaration, the evidence for HIV as the cause of AIDS is overwhelming. Drugs such as AZT can be used relatively cheaply and effectively to reduce mother-to-child transmission of this virus. (Readers interested in further details about HIV as the cause of AIDS can see <http://www.niaid.nih.gov/factsheets/evidhiv.htm>. Those wishing to assess the opponents' case might visit <http://www.duesberg.com>.)

Unlike some websites disputing the evidence that HIV is the cause of AIDS, to which anyone can apparently add their support regardless of their expertise, the Durban Declaration is signed only by scientists and representatives of organizations involved in research and treatment of AIDS. That is not the reason *Nature* endorses their position, however — indeed, *Nature* takes pleasure in publishing papers that convincingly overturn widely accepted hypotheses. It is, rather, on the compelling basis of the science that we urge that the declaration's message be acted upon by President Mbeki and his government. ■





Pooled resources
Japan delves into the
genomes of humans
and chimps
p4



Health problems
Technology transfer
at NIH comes under
fire in United States
p5



In sequence
Germany banks on
genomics for future
research
p6



Suicide mission
NAS backs proposed
fate of Jupiter
space craft
p8

African scientists join colleagues in affirming HIV's role in AIDS...

AP

Michael Cherry, Johannesburg

About 400 independent African scientists, including the health minister of Swaziland, Phetsile Dlamini, are among those who have signed an international declaration stating that HIV causes AIDS. The document also condemns revisionist theories that the virus is not the cause of the disease.

More than 5,000 scientists and physicians worldwide — none of whom are employees of pharmaceutical companies — have signed what is being called the Durban Declaration (see page 15). It is being issued on the eve of the 13th international AIDS conference, which starts in Durban on Sunday.

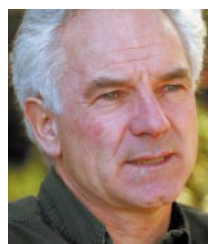
The declaration, organized by a 266-member committee, states that the evidence that AIDS is caused by HIV is “clear-cut, exhaustive and unambiguous”. It adds: “It is unfortunate that a few vocal people continue to deny the evidence. This position will cost countless lives.”

Dissident scientists sitting on the advisory panel set up by President Thabo Mbeki to address the nature of AIDS and its links to HIV (see *Nature* 405, 105; 2000) have criticized fellow panel members who signed the declaration. The dissidents claimed that the signatories, including Malegapuru Makgoba, president of South Africa's Medical Research Council (MRC), were prejudging the issues being considered by the panel.

But South African minister of health, Manto Tshabalala-Msimang, effectively neutralized the possibility of the dissidents using the panel as a forum to oppose the declaration. Speaking at the opening of a two-day meeting of the panel in Johannesburg earlier this week (see right), she declared: “We uphold the right of every scientist to append their signatures to a document which articulates a scientific point of view to which they subscribe.”

Prominent dissident Peter Duesberg, professor of molecular biology at the University of California at Berkeley and a member of a four-person task force set up by the Mbeki panel to design experiments to test the link between HIV and AIDS, says that “scientific pogroms” are common.

“In this case, the majority has persuaded



Duesberg (left) dismisses ‘scientific pogrom’ but Tshabalala-Msimang is more conciliatory.



the world's oldest scientific journal to publish a declaration designed to intimidate scientific minorities from questioning a hypothesis that has yet to cure a single AIDS patient,” he said in Johannesburg this week.

But Makgoba points out that the declaration “is about the history and science of the HIV/AIDS epidemic from scientists who have direct experience in managing and researching it”.

Other panel members who have signed the declaration include Hoosen Coovadia, convener of the Durban conference; Helene Gayle of the Centers for Disease Control and Prevention in Atlanta, Georgia; Salim Karim, director of HIV prevention and vaccine research for South Africa's MRC; and

physicians James McIntyre and Glenda Gray of the Chris Hani Baragwanath Hospital in Johannesburg.

Makgoba describes the attempt to make the presidential advisory panel a platform for the dissidents to respond to the declaration as “opportunism”. He adds that “the crunch has come for them, as experimental evidence in support of their view is being demanded of them — something they have never been able to supply”.

Parks Mankahlana, Mbeki's spokesperson, commented to the Johannesburg newspaper *The Star* that the president “respects the rights of people to issue declarations and publish them, but we must be careful that we don't turn the Durban conference into an Mbeki-bashing bazaar”.

In an interview published in the *Sunday Independent* last weekend, three South African cabinet ministers — including Tshabalala-Msimang and arts, culture, science and technology minister Ben Ngubane — defended Mbeki against accusations that his discussions with scientists disputing the link between HIV and AIDS implied that he himself denied the connection. “Simply put, the president has never stated that HIV does not cause AIDS,” the ministers said. ■

...but advisory panel remains divided

Five days before the start of the 13th international AIDS conference in Durban on Sunday, an advisory panel on AIDS set up by South African President Thabo Mbeki met there amid optimism that the government might be prepared to modify its position.

The enlargement of the panel from 36 members to 50, with the new members largely comprising mainstream South African and Ugandan AIDS researchers, has led to speculation that the South African government might be about to review its policy on AIDS.

The non-dissident panel

members are drawing up a proposed government policy on the prevention and treatment of AIDS.

Members of the dissident and non-dissident camps on the panel are believed to have worked largely independently of each other. The dissident faction has continued to argue for policies based on their claims that AIDS is not contagious and not sexually transmitted.

But further evidence that HIV causes AIDS came from a study by panellists James McIntyre and Glenda Gray on a cohort of babies born to HIV-positive mothers in the township of Soweto who were

not treated with antiretroviral drugs. Infant mortality rates after one year were 17/1,000 in babies who were HIV negative after delivery, compared with 326/1,000 in those who were HIV positive.

As *Nature* went to press, there seemed little chance of a consensus emerging between the dissident and non-dissident panellists. National Research Foundation president Khotso Mokhele, as head of the panel's secretariat, is faced with the task of compiling a report for Mbeki before his opening address to the Durban meeting.

M. C.

Royal Society warns on hormone disrupters

Natasha Loder, London

Britain's science academy is calling for worldwide action to deal with chemicals that may disrupt the body's hormonal functions.

In a report published last week, the Royal Society calls for national and international coordination to deal with the dangers it claims are posed to humans and wildlife by endocrine-disrupting chemicals (EDCs) — substances that are thought to mimic or block hormones, in doses too small to trigger a conventional toxic response.

The report, called *Endocrine-disrupting Chemicals*, was produced by a working group led by the society's vice-president Patrick Bateson, professor of ethology at the University of Cambridge. It says there is strong evidence to link EDC exposure to effects on some organisms. It recommends minimizing human exposure to EDCs — with pregnant women being strongly discouraged from contact with EDCs, such as plasticizers and insecticides.

But research is being hampered by lack of standardization. A US study of 15,000 chemicals, for example, has been unable to begin a planned high-throughput screening programme because researchers have been unable to develop the right assays.

"There are no ways to test these chemicals, no standardized screens or assays, even in one nation," says Theo Colborn, an expert on the endocrine-disrupter hypothesis who works at the World Wildlife Fund in Washington, DC.

Colborn says that new research has recently become available on prenatal exposure to the industrial chemicals polychlorinated biphenyls (PCBs) and co-contaminants, which shows a correlation with reduced IQ, behavioural problems and adverse effects on the immune system.

The British report goes further than one last year from the US National Research Council (NRC), which left an open verdict on the issue of public health risks (see *Nature* 400, 607; 1999).

Some scientists remain sceptical of the endocrine-disrupter hypothesis. Stephen Safe, for example, deputy director of the Center for Environmental and Rural Health at Texas A&M University and a member of the NRC panel, says that the recent evidence about PCBs is not clear-cut. Safe says that, although he does not necessarily disagree with the endocrine-disrupter hypothesis, he considers that it needs more work, and is not convinced by the evidence so far.

"Do EDCs cause some problems? They have and probably will, but the general hysteria is unwarranted," says Safe. He blames what he calls "synthetic chemical chemophobia" for current public reaction.

In contrast, Bateson says the issue is a

"cause for more worry" than genetic modification. He thinks there is a case for making people more aware of the possible dangers, and is arguing for a national body to oversee work on the issue. The society wants to ensure that "sound policies" are developed and warns policy-makers that they must appreciate that the concerns of the public already have some foundation.

The Royal Society report coincides with growing international concern about the harmful effects of EDCs.

In April a meeting of the environment ministers of the G8 group of industrialized countries signed a communiqué stating that the risks posed by hazardous chemical substances were one of the greatest concerns

expressed by the people of their countries. The ministers called for a "furtherance of knowledge acquisition on endocrine disrupters through jointly planned and implemented projects and international information sharing".

Colborn is currently touring the world in a bid to raise money to set up an independent international body, which would be funded by industry and government and would co-ordinate research and expertise.

Earlier this year the UK's Environment Agency published a strategy for reducing potential EDCs. But Bateson thinks European bodies — including those in Britain — are "bitty and uncoordinated, with lots of activities all over the place". ■

Study compares chimps and people

Robert Triendl, Tokyo

Scientists at two major Japanese research facilities are finalizing plans for a joint project aimed at comparing the genome of humans with that of chimpanzees.

It will be run by the Brain Science Institute (BSI) and the Genomic Sciences Center (GSC), both of which are part of the Institute of Physical and Chemical Research (RIKEN), Japan's foremost institution for basic science.

The project will be directed by Yoshiyuki Sakaki, who holds joint appointments at the Institute of Medical Science at the University of Tokyo and at the GSC. It will initially concentrate on the comparison of gene clusters expressed in the region within the brain responsible for speech and language.

"It is intellectual ability and especially language that distinguishes humans," says Sakaki. "This project will help us to identify genes that are distinctive of humans and linked to the intellectual capacity of humans."

The project will initially focus on genetic comparison, but a behavioural component may be added. Japan has a rich tradition of ethological research on primates.

Although work on the chimpanzee genome has barely begun, scientists think that experience in sequencing the human genome can be replicated without much problem in the case of chimpanzees.

As structural genomics and research on the mouse genome — areas where Japan has made early commitments — enter the mainstream of genome research, it is hoped that the new project will allow Japanese scientists to pursue an original field of research. "I believe this is an area where Japanese scientists can make a clear difference," says Sakaki.



Two of a kind? A Japanese project could reveal exactly what humans and chimpanzees share.

Although funds have not yet been committed, the Science and Technology Agency (STA), which oversees RIKEN, is said to be eager to promote the project for another reason.

At a recent meeting of the international advisory board of the GSC, board members expressed concern that the centre's ample facilities and resources had not been made available to outside scientists.

Both the GSC and the BSI are the result of a shift towards priority-based funding that has gained momentum since the early 1990s. So far, it has not proved easy to integrate the new centres into the diverse landscape of Japanese public research.

According to its director, Masao Ito, the BSI provides small seed funding grants for cooperation with scientists outside the centre. But he acknowledges that these are fairly small, and that it has been difficult to convince officials at the STA to upgrade resources for collaborative research. ■

KAREN HUNTT MASON/CORBIS

NIH urged to cap profits made on publicly funded research

Colin Macilwain, Washington

The US National Institutes of Health (NIH) is caught in the crossfire over one of the most contentious issues in the US presidential campaign: the high cost of therapeutic drugs.

In particular, the medical agency is under political pressure to cap or recover profits on drugs whose development was based on publicly funded research. But this is being opposed by the pharmaceutical and biotechnology industry, and also by US universities and medical colleges, which benefit substantially from the current arrangement.

Last week, intensive lobbying by the pharmaceutical industry helped to defeat two amendments to the NIH's Senate funding bill. These would have forced radical changes in the rules for the transfer of technology resulting from publicly funded research from universities to corporations.

One amendment, proposed by Senator Ron Wyden (Democrat, Oregon) but later withdrawn, would have required universities to repay grants to the NIH if research led to a commercial product. And drug companies would have had to pay a 1% levy to the NIH on drugs with annual sales of \$500 million.

Under the current system, implemented largely through the landmark 1980 Bayh-Dole Act, companies often pay licensing fees to the institution that patented an innovation. But neither the companies nor the institution have to pay the government anything.

The Bayh-Dole Act was intended to allow corporations freedom to patent and profit from publicly supported university research without entering into complex legal agreements with the government. The act is credited by some economists as key to the success of the US biotechnology industry, which has grown much faster than its counterparts abroad.

But the high prices charged by the drug industry for drugs partially developed with public funds do not sit well with Congress.

"The thing that is providing the heat on this issue is the intense political jockeying over prescription drug benefits," says David Korn, vice-president for research at the Association of American Medical Colleges.

Last week, an amendment from Wyden was added to the Labor, Education, Health and Human Services appropriations bill. It requires the NIH director to report to the Senate "a proposal to require a reasonable rate of return on both intramural and extramural research by March 31st, 2001". But this addition came after Wyden had withdrawn his initial wording.

Another amendment, from Senator Paul Wellstone (Democrat, Minnesota), would have prohibited licensing deals unless the



Sanders: claims Americans are currently paying twice for life-saving drugs.

drug companies involved reached "reasonable pricing" agreements with the government. It was rejected by the Senate last Friday.

The sentiments behind this amendment are already supported in the House of Representatives. Last month, the House approved an amendment to its NIH funding bill, proposed by Bernie Sanders (Independent, Vermont), to reimpose "reasonable pricing".

John Porter (Republican, Illinois), chair of the House Labor, Education Health and Human Services appropriations subcommittee, says the amendment will have little effect. Although it prohibits NIH funding for licensing agreements, these are normally funded directly by universities.

But the amendment's easy passage indicates the strength of feeling on the issue. Many in Congress find it hard to argue with Sanders' line that "Americans must pay twice for life-saving drugs, first as taxpayers to develop the drug and then as consumers to pad pharmaceutical profits". In a newspaper interview last week, Al Gore, the Democratic nominee in this year's presidential election, said that he supported the Sanders amendment.

Porter, who opposed the amendment, says the NIH already demands a return when it collaborates with industry on research directed at a particular disease. But he says the argument now is about demanding returns from basic research, whose contribution to drug development is hard to measure.

The NIH sought to impose "reasonable pricing" of drugs when Bernadine Healy was its director, but her successor Harold Varmus abandoned the idea in 1995.

But a spokesman for the Association of American Universities says it is "concerned about opening up the Bayh-Dole Act through the appropriations process, without thorough hearings and careful deliberations".

The failure of the Wellstone amendment lifts that threat for now. But Bayh-Dole still looks vulnerable, says Korn. "In a presidential election year, anything could happen."

Indian space agency sets its sights on a mission to the Moon

K. S. Jayaraman, Bangalore

The Indian Space Research Organisation (ISRO) is to submit a feasibility report to the government seeking permission and funding for a scientific mission to the Moon. The ISRO says it can launch the mission by 2005 if it gets the go-ahead.

The report, due to be delivered in about three months, is being prepared by an ISRO team assembled to look into the feasibility, cost and benefits of the project.

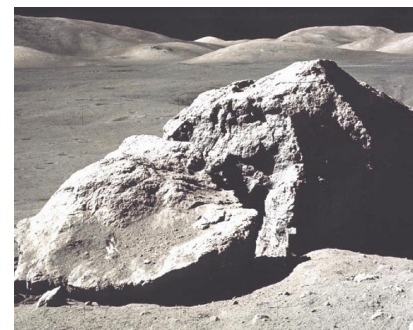
S. Rangarajan, mission coordinator and ISRO director for satellite communications, says the mission will cost about US\$90 million — roughly a fifth of the ISRO's budget for 1999. He says it will be "a one-shot affair, not a continuing moon programme".

The ISRO is considering modifying India's Polar Satellite Launch Vehicle to send a lunar orbiter carrying sensors and high-resolution cameras. "The orbiter will be a lighter version of the remote-sensing satellites we have been building for over a decade," says P. S. Goel, director of the ISRO Satellite Centre.

According to Rangarajan, the mission has already been endorsed by India's scientific community. And ISRO chairman Krishnaswami Kasturirangan says the mission would "provide an impetus to science in India" and serve as a test-bed for future space missions.

But not all scientists are so keen. H. S. Mukunda, chairman of the aerospace engineering department at the Indian Institute of Science in Bangalore, says the country is unlikely to benefit from repeating what others did 30 years ago.

The ISRO has listed several possible scientific objectives, including detailed mapping, studying the distribution of rare elements, and analysing the surface composition. It has also invited proposals from the research community.



Space odyssey: India hopes to send an orbiter to investigate the Moon.

Good news for German genome research...

Alison Abbott, Munich

Last week's announcement of the draft sequence of the human genome has prompted politicians in both Germany and Italy to publish papers on national programmes of genome research that had been gathering dust for many months.

German scientists have welcomed the federal research ministry's plan for a significant increase in support for such research. But their Italian colleagues are dismayed by support for a bill introduced into the country's senate creating a politically controlled oversight committee that would be responsible for all aspects of genome research (see story below).

Germany came to the genomics game late; the German Human Genome Project (DHGP) was launched, with limited public funds, in only 1996. But research minister Edelgard Bulmahn, who was originally sceptical about the importance of genomics, has been persuaded otherwise, primarily through the persistent efforts of her secretary of state Wolf-Michael Catenhusen.

As a result, the ministry this week published a detailed financing plan for the next four years which envisages an overall 50% increase in project and institutional money. A main aim is to concentrate and coordinate resources. Project money for the DHGP will increase from DM44 million (US\$21 million) to DM66 million.

Germany's resource centres, which provide genomics services such as the creation, archiving and distribution of clone libraries for research groups, were set up with temporary funding, and they will now become permanent bodies.

A series of competitions will be launched



Help at hand: resource centres will be permanent.

to support regional collaboration between industry and academic researchers. One will select three 'competence' centres, or networks, which will receive money from a DM100 million fund; this must be matched by money from other sources. The DHGP will receive a further DM50 million to distribute between these centres.

A major effort in microorganism genomics will be launched using DM40 million to support research networks among universities. Money for existing programmes in plant genomics and bioinformatics will be increased by around two-thirds. And a DM134 million fund will support proteomics research, the analysis of protein profiles.

"We're happy," says Rudi Balling, head of mammalian genetics at the Munich-based GSF national research centre and a member of the DHGP scientific coordinating committee. "Both the money and the psychology are right." Balling points out that it is the first time that the Social Democrats, the major coalition partner in the government, have made a

strong commitment to genomics. "They have always been rather vague, but now it seems they are really 'outing' themselves."

"The size of the announced increase has taken us by surprise," says Detlev Ganten, president of the Helmholtz Society, the umbrella group for Germany's 16 national research centres. "The research ministry's plans are very close to what the scientific community has been lobbying for."

Ganten was one of 30 key German scientists who signed a memorandum circulated last week calling for a more concerted effort in support of genome research from both the public and private sectors. The ministry's move will encourage private investment in genomics, he says.

Catenhusen, who has lobbied relentlessly for genomics for more than a decade, is basking in the positive response. "As a politician, I can tell you this programme is an extraordinary success, considering we started from virtually nothing."

But not all are happy. André Rosenthal, professor of molecular biology at the University of Jena and managing director of the Berlin-based genomics company MetaGen, says that, even with the increase, Germany will still be spending only a third of what the Deutsche Forschungsgemeinschaft, Germany's university granting agency, said two years ago was necessary to catch up with the United States.

Moreover, he says, the competence networks and centres are predestined to be based in Berlin, Munich and Heidelberg, since they are intended to build on existing concentrations of research. "It is not real competition — smaller centres will not get a look in," he complains. ■

...but proposed panel may hamper Italian gene researchers

Significant aspects of genomics research in Italy will be controlled by a three-member panel of experts nominated by parliament, if a private member's bill is successful.

The bill was approved last week by the culture and education committee of the Italian Senate, but still has some hurdles to clear before becoming law.

The panel would have a seven-year mandate. Based on advice from expert committees, it would make final decisions about what type of basic and applied research could be done, taking into account both bioethical and safety issues, and with responsibility for ensuring 'a balance of ethical positions' within the research community.

The panel, whose creation was initially proposed by senator Valentino Martelli, would not distribute research money, but would have the power to block funds distributed by other agencies for research it considered unsuitable. Ultimately, it would have the power to close laboratories conducting research outside its guidelines.

The move has wide support across different political parties. But many scientists see it as a clumsy response to ethical concerns about human genome research, claiming that it is ill-thought-out and makes no reference to existing national bioethical and scientific advisory committees. Research minister Ortensio Zecchino

has also expressed reservations, but a private member's bill does not require government support to be debated.

Although the bill is fast-tracked and has considerable cross-party support, it may not be approved in the current government's legislative period, not least because of its high proposed running costs of IL40 billion (US\$19.6 billion).

Its lack of reference to existing structures highlights the poor coordination of genomics research in Italy, which was among the first countries after the United States to institute a genomics-research programme. The programme, an initiative of the Italian national research

council (CNR), was led by Nobel laureate Renato Dulbecco. It was launched with only IL2 billion, and funding stopped in 1996 when the CNR came under political pressure to reform.

The CNR is now launching a new genomics-research programme intended to prepare the ground for a more extensive initiative. Its IL1.9 billion budget is less than one-twentieth the annual running costs proposed for the planned controlling panel.

Zecchino responded to last week's announcement of the human genome draft sequence by pledging to create a high-level school for molecular medicine, with sites in Milan and Naples. But he gave no details. **A.A.**

Self-policing backed for research on humans

Colin Macilwain, Washington

The presidents of America's top research universities have come up with a set of proposals for tightening up the policing of research on human subjects.

The plan, published in Washington last week by the Association of American Universities (AAU), addresses mounting public anxiety about the safety of individuals taking part in such trials. Levels of concern have been high since the death last September of teenager Jesse Gelsinger, who died while undergoing gene-therapy trials at the University of Pennsylvania (see *Nature* 401, 517; 1999).

"University presidents, such as myself, have to be more involved" in the supervision of human-subject research, says Dennis Smith, president of the University of Nebraska at Lincoln and co-chair of the AAU task force behind the plan. University presidents would welcome greater involvement in regulating research, adds Nils Hasselmo, president of the AAU, as the problems that result from lax regulation "end up on the presidents' desks".

The AAU, which represents the presidents of 59 top US and two Canadian research institutions, says its proposals would require considerable investment from its members. But its plan goes further than the regulations being implemented at the universities by the federal government.

The scheme would involve university presidents more deeply in monitoring clinical research. It would commit resources for better training of the researchers involved in

human-subject research, and strengthen the institutional review boards (IRBs) that review clinical trials. "Money needs to be made available to give the IRBs sufficient resources to do the job," says Smith. "Another priority is to put more members of the public on IRBs."

The AAU's plan would also establish a system of 'voluntary accreditation', similar to that in effect for animal research, to reassure the public that research subjects are protected.

Hasselmo says that most university presidents support the legislation proposed by Representative Diana DeGette (Democrat, Colorado), which aims to strengthen the fragmented IRB network. The Association of American Medical Colleges (AAMC), repre-

senting the medical schools, has already endorsed the bill. The AAU has yet to adopt a formal position.

Some patient advocacy groups say that self-regulation of human-subject research has already failed, most notably in Gelsinger's case. They want independent, external regulation, pointing out that neither the AAU nor the legislation before Congress provide this.

Under both DeGette's bill and the AAU's plan, institutions conducting human-subject research would register voluntarily for accreditation with an independent body. The Boston-based non-profit organization Public Responsibility in Medicine and Research (PRIM&R), which trains researchers and IRB members in human-subject protection, is preparing itself to fulfil this role, with the blessing of the AAMC and the AAU.

But Vera Sharav, president of Citizens for Responsible Care and Research, a New York-based group advocating research subjects' rights, dismisses PRIM&R as "an organization that represents the IRBs". She ridicules the AAU for proposing "independent self-monitoring" at institutions. "Self-monitoring cannot be independent," she says. The AAU says the phrase refers to monitoring by the universities, independent of the IRBs, and that the accreditation system will provide additional, external monitoring.

A spokesman for DeGette says the bill has strong bipartisan support, and that it should hopefully be the subject of Congressional hearings before the end of the month. ■



Austrian physics meeting passes without boycott

Quirin Schiermeier

The high attendance at a physics meeting in Vienna last week — as well as the continued refereeing of Austrian research proposals by foreign scientists — suggests that fears of a boycott of Austria by the international scientific community may be unfounded.

But several hundred of those attending the seventh European Particle Accelerator Conference (EPAC) signed a letter to Austrian Chancellor Wolfgang Schüssel saying that their participation "should not be misconstrued as an indication of support for the present Austrian government".

Earlier this year, the US Department of Energy and the American Physical Society discussed a possible boycott in response to the inclusion in Austria's coalition government of the far-right Austrian Freedom Party.

Although the United States stepped back from an official boycott, some laboratories, such as the Brookhaven National

Laboratory, and many scientists both inside and outside the United States pledged to boycott the meeting, leading to fears that it could face serious economic difficulties.

But many apparently changed their minds at the last minute, possibly in response to calls from Austria's science agencies for support from the international scientific community (see *Nature* 403, 691; 2000). The EPAC was attended by 750 physicists, including 103 from the United States. Although 10% fewer than initially expected, this was more than attended the 1998 EPAC in Stockholm.

The letter to the Austrian government was drafted by scientists in the United States and at CERN, the European Laboratory for Particle Physics. Signatories included Andrew Sessler, former president of the American Physical Society, and Yuri Orlov, the honorary chairman of the International Helsinki Federation for Human Rights.

Representatives of the Austrian government were excluded from the meeting

on the advice of Steve Myers of CERN, chairman of EPAC's international organizing committee.

"I personally think that at least the Austrian science minister, Elisabeth Gehrer, should have been invited to give a message of greeting," says Meinhard Regler, vice-director of the Vienna-based Institute for High Energy Physics and chairman of the local organizing committee. "But I agree that the presence of an Austrian minister could have easily led to embarrassing situations."

Arnold Schmidt, president of the Austrian Science Fund, Austria's main research-funding agency, says the political situation has not caused serious problems for Austrian scientists. Responses to only four out of 2,500 grant applications sent to referees outside Austria mentioned the Europe-wide restrictions of diplomatic relations with Austria, and there has been no significant reduction in the return rate of reviews, he says. ■

New law threatens to undermine genetics in New Zealand

Peter Pockley

Researchers and government officials in New Zealand are pinning their hopes on a royal commission to end their struggle over a strict law governing use of genetically modified organisms (GMOs) in research (see *Nature* 404, 914; 2000).

But the Royal Commission on Genetic Modification might not report for up to two years. During that time, says Clive Ronson, the biological safety officer at the University of Otago, "New Zealand scientists have to face retrograde legislation that is based on organisms rather than projects and undermines our competitiveness".

Last April, government officials suspended the authority of all institutions' internal committees to license GM research. The move followed investigations by the Environmental Risk Management Authority (ERMA) of experiments that had not been approved under the Hazardous Substances and New Organisms (HSNO) Act.

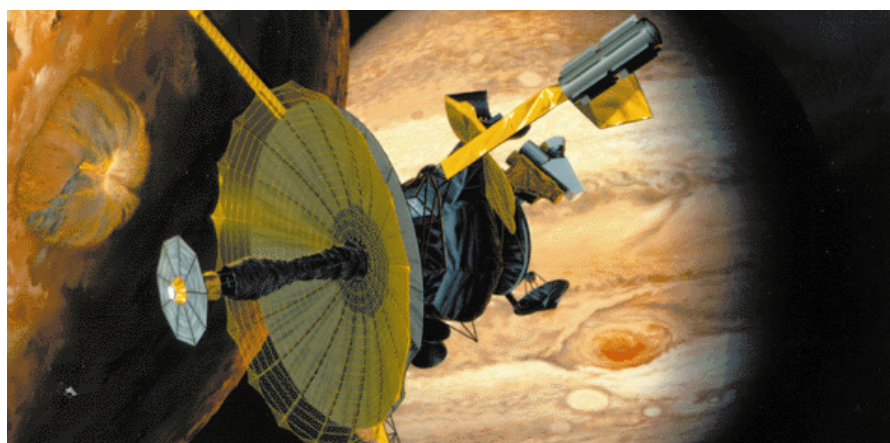
Researchers are now protesting at the fact that ERMA classes every modification to an organism, however minor, as a new organism, requiring a lengthy and costly application for approval (see page 13). As a result, hundreds of projects have been placed on hold, awaiting approval.

Although few have made public statements, biochemist Jon Hickford of Lincoln University near Christchurch recently told the city's daily newspaper, *The Press*, that ERMA had failed to explain to the public the "low risk" posed by routine modification of plasmids in bacterial strains. He said ERMA should concentrate on the larger challenges posed to the biosecurity of New Zealand's agricultural economy by animals and insects entering the country.

Bas Walker, ERMA's chief executive, responded by accusing Hickford of "trying to shoot the messenger because the message is unpalatable".

Last week, the University of Auckland announced that it would have to devote up to NZ\$1 million (US\$468,000) of its NZ\$15 million GMO-research funding to meeting what it calls "inappropriate requirements" imposed by the HSNO Act.

Walker revealed that, after a nationwide check by ERMA following the first breach, 27 facilities throughout New Zealand had reported 1,065 GM experiments, of which 152 had not received approval. ■



Life saver? The spacecraft Galileo will plunge into Jupiter's atmosphere to protect Europa (left).

Galileo set for suicide flight

Heather McCabe

A US National Academy of Sciences panel last week endorsed plans to send NASA's Galileo spacecraft on a suicide mission into Jupiter's atmosphere. The move is a bid to avoid contaminating the planet's potentially life-harboring moons, in particular Europa.

At the same time, the Committee on Planetary and Lunar Exploration, part of the academy's Space Studies Board, agreed to get as much science as possible out of the spacecraft before its navigation systems fail. Galileo was originally launched in 1989 and has been orbiting Jupiter since 1995.

The spacecraft is scheduled to complete a series of fly-bys of Io, one of Jupiter's moons, delaying the craft's destruction by a year. But

this is not thought likely to increase the chance of an accidental crash into Europa. Some believe that this moon has a liquid ocean under its ice crust that could contain simple life-forms. Any terrestrial organisms still alive on Galileo could potentially contaminate the moon.

"The probability of total loss of control during this extra year is relatively small," concludes the report. "Moreover, the chances of total failure can be mitigated by prudent monitoring of the spacecraft's health and by a commitment on the part of NASA to retarget Galileo onto a Jupiter-bound trajectory following the loss of redundancy in any major command and control subsystem." ■

Astronomers fight to save telescope

Rex Dalton, San Diego

Astronomers are fighting to save the 12-Meter Telescope at Arizona's Kitt Peak from closure at the end of July. The National Radio Astronomy Observatory hopes to save about \$2 million by mothballing the telescope — the only millimetre-wavelength facility available to all US astronomers.

The observatory operates the telescope for a consortium of universities (see *Nature* 404, 7; 2000). But the University of Arizona in Tucson and the University of Massachusetts at Amherst joined forces with dozens of astronomers late last month to develop a proposal to keep it operational.

Millimetre-wavelength telescopes are used to examine the early phases of star formations in supercold molecular clouds. For instance, researchers using the 12-Meter Telescope recently discovered deuterium (a form of hydrogen) in the Milky Way (Ludowich, D. A. *et al.* *Nature* 405, 1025–1027; 2000).

Astronomers from Arizona and Massachusetts met last week with the National Science Foundation (NSF), which owns the facility. "The astronomy community is working very diligently to come up with a solution," says James Breckinridge, an NSF astronomy programme manager.

Later this month, the universities plan to submit an interim proposal to the NSF to operate the telescope with the agency. To begin operations on 1 August, the two universities have secured a \$150,000 donation from Research Corp., a Tucson-based non-profit-making corporation.

The NSF will decide later this month what will happen on 31 July when the telescope is due to shut down. The astronomers are also writing a three-year research plan, to be sent to the NSF later this year.

A joint US–Mexican project is under way to build a millimetre-wavelength telescope with a 50-metre dish in Mexico. But this will not come into operation until 2003. ■

Large funding boost keeps NIH on course for doubling of budget

Washington The US Senate last week passed an appropriations bill that would increase next year's budget for the National Institutes of Health (NIH) by \$2.7 billion to over \$20 billion. The increase, if passed into law, would keep the NIH on track to double its budget over five years from 1998.

The Senate proposal has to be reconciled, later this summer, with proposals from the Clinton administration and the House of Representatives to raise the NIH budget by only \$1 billion. But the agency is now strongly positioned to obtain the massive increase approved by the Senate.

In a less auspicious but equally preliminary development, the House appropriations subcommittee for energy proposed to cut funding for basic energy sciences by \$224 million below Clinton's request. The move threatens the continued construction of the Spallation Neutron Source at the Oak Ridge National Laboratory in Tennessee. The Senate has yet to consider funding for energy research.

Glaxo chairman takes over at Imperial College

London Glaxo Wellcome has announced that its chairman, Sir Richard Sykes, has been appointed the next the rector of Imperial College of Science, Technology and Medicine in London. He will take up the role in January 2001. The announcement is seen as a strong indication that merger plans between Glaxo Wellcome and SmithKline Beecham are proceeding smoothly — despite the announcement last month that they are being scrutinized by the Federal Trade Commission (FTC).

Sykes will become non-executive chairman of Glaxo SmithKline plc when the merger is complete. If FTC approval is received, the deal must also be approved by shareholders. Documents outlining the deal will be passed to shareholders this month.

Australia buys access to genome from Celera

Sydney Australia is claiming to be the first country to sign a deal with Celera Genomics to allow researchers access to the database of the human genome, as well as to those of the mouse and the fruitfly *Drosophila*. The deal was struck with Australia's National Health and Medical Research Council. Warwick Anderson, chairman of the council's research committee, says it is "extremely favourable" for Australia.

Publicly funded institutions will be among the first in the world to work with

Celera's information technology. Although Celera has not disclosed financial terms, each institution in Australia will pay a minimum annual licence fee of about A\$6,000 (US\$3,500). As well as medical researchers, those funded through the Australian Research Council (mostly in universities) and the Commonwealth Scientific and Industrial Research Organisation will gain access.

University of Washington pathologist killed by student

San Diego A Chinese researcher shot and killed a leading pathologist at the University of Washington last week, then committed suicide on the Seattle campus. The victim was Rodger Haggitt, head of anatomical pathology at the university's medical centre and a prominent gastroenterology researcher.

Jian Chen, who held a medical degree from Shanghai Second Medical University and a biology doctorate from Barrington University in Iowa, carried out the attack after being dropped from a physician training programme because of inadequate performance and poor English.

Spanish university sprouts botany centre

Barcelona The University of Valencia last week launched its botany research centre, which will focus on conservation and biodiversity. Mostly funded by the European Commission, the new building cost Ptas 800 million (US\$ 4.5 million). It will house 30 researchers working on projects including biosystematics, molecular biology, bioclimatology, geobotanics and phytosociology.

The centre will feature a herbarium containing 300,000 plant samples from locations across the western Mediterranean. The collection has been built up over many years by researchers at the university and by exchanges with herbaria at other European institutions. The research centre will also include a unique germplasm bank, which aims to guarantee the long-term survival of threatened or rare flora in the region.

London's Science Museum launches education wing

London The Science Museum in London last week launched the "most ambitious" project in its history. The new £50 million (US\$75 million) Wellcome Wing, sponsored by the Wellcome Trust, contains a set of exhibitions on key scientific topics including genetics, digital technology, biomedicine and artificial intelligence.

The new wing is also designed to prompt debate about wider ethical, social and political issues associated with science. A



Getting a feel for patterns at the Wellcome Wing.

hands-on section known as the 'pattern pod' (see above) has been designed to introduce young children to the role of patterns in contemporary science.

German grant money goes to the few

Munich More than half of the grant money distributed between 1996 and 1998 by the Deutsche Forschungsgemeinschaft (DFG), Germany's main research funding agency, went to only 20 of Germany's roughly 300 universities. According to the 1996–1998 breakdown of successful grant applications, presented last week at the DFG's annual meeting, the Ludwig Maximilian University of Munich topped the list with grants worth a total of DM196 million (US\$95 million), followed by the Technical University of Aachen (DM188 million).

The Technical University of Munich (DM181 million) came third, followed by the universities of Heidelberg, Stuttgart, Tübingen, Hamburg, Erlangen–Nürnberg, Berlin (Humboldt University) and Würzburg. Universities in eastern Germany have won considerable ground since the 1991–1995 ranking, with the Humboldt University leaping 21 positions up the ranks into the top ten.

► <http://www.dfg.de/berichtswesen/ranking.html>

California loses security responsibility after lapses

San Diego The University of California is to be stripped of security responsibilities at the two national weapons laboratories it manages, after a series of security lapses. The US Department of Energy announced last week that it is renegotiating the university's long-term contract to manage the Los Alamos National Laboratory in New Mexico and the Lawrence Livermore National Laboratory east of San Francisco. The university will continue to oversee the scientific enterprises at the labs, as it has done since shortly after the Second World War.

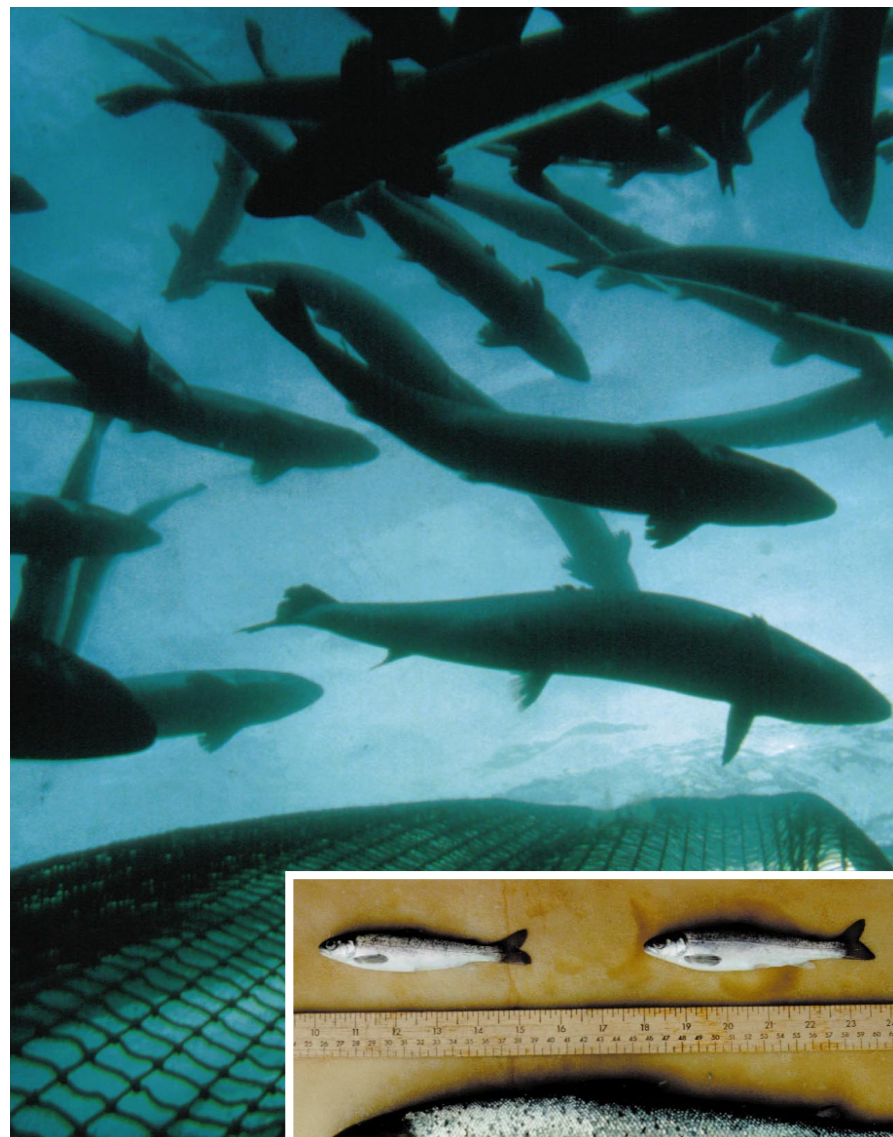
Will souped up salmon sink or swim?

A company in Massachusetts is seeking permission to market salmon genetically modified to grow faster than normal. Tony Reichhardt explores the potential ecological risks, should the fish escape from salmon farms.

A few years ago, when Garth Fletcher's office telephone rang, it was usually another scientist wanting to talk about aquaculture or fish genetics. Nowadays, it is just as likely to be a news reporter asking barbed questions about 'Frankenfish'. "We're getting hit every day in the press," he says.

Fletcher is president of the Canadian arm of Aqua Bounty Farms, a company based in Waltham, Massachusetts, that hopes to bring genetically modified (GM) salmon to the dinner plates of North America. At the company's experimental hatchery on Canada's Prince Edward Island, its aquaculturists are raising Atlantic salmon (*Salmo salar*) modified to carry a growth-hormone gene from the Pacific chinook salmon (*Oncorhynchus tshawytscha*), which is hooked to a powerful promoter sequence. This boosts the fishes' growth rate, so that they reach market size quicker¹.

Aqua Bounty Farms, formerly a subsidiary of the company A/F Protein, has



Scaled up: the effects of growth-hormone genes (right) may find use in fish farms (above).



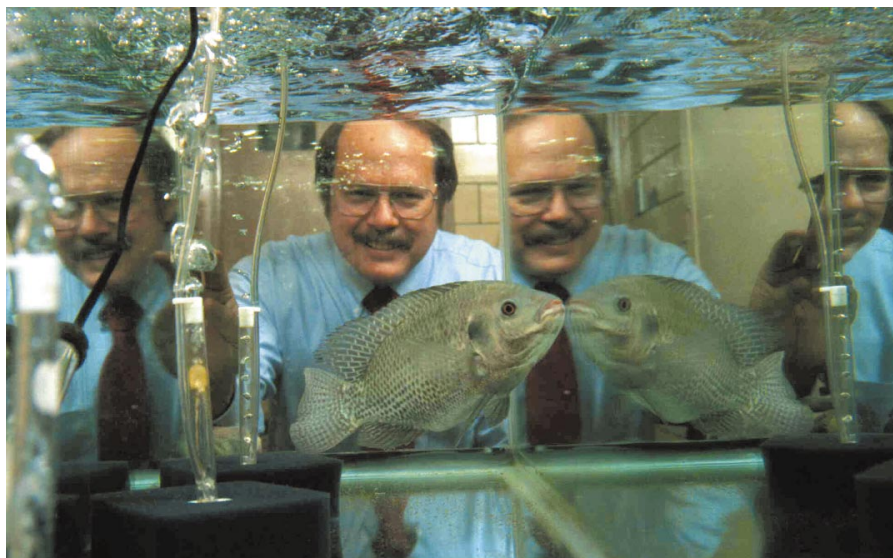
applied to the US Food and Drug Administration (FDA) for permission to market its salmon. And ever since this application was picked up by the media, the company has been plunged into the thick of the controversy surrounding GM foods — which, after a slow start compared with the furore in Europe, is now beginning to register in North America. Some ecologists have even warned that transgenic salmon could wipe out natural populations of related fish should they escape into the wild.

So far, the only things that have been wiped out are the business plans of two companies that licensed A/F Protein's gene-insertion technology in the 1990s. Both Otter Ferry Salmon in Scotland and the New Zealand King Salmon Company scrapped their GM salmon research after unfavourable publicity. But with Aqua Bounty Farms still pressing ahead, ecologists warn that the current state of scientific knowledge is

inadequate to provide a full assessment of the risks posed by the company's fish. "There's just so much speculation compared to the amount of data," says Robert Devlin, a researcher with Canada's Department of Fisheries and Oceans, based in West Vancouver. Much of the research only started in the past decade. Because it can take 10 years to produce a stable line of transgenic salmon, says Devlin, the dearth of experimental studies is hardly surprising.

Research on transgenic strains is currently under way for some 35 species of fish worldwide, including Pacific salmon such as the chinook and coho (*Oncorhynchus kisutch*), various other members of the salmonid family, and other economically important fish including catfish and tilapia. Most of the work is being done by commercial fisheries and involves growth-hormone genes.

The one certainty is that conventionally farmed salmon, typically raised in netted pens



Trojan genes: Muir's research has raised fears that wild salmon may be decimated by GM fish.

in shallow coastal waters, will escape. Where there are salmon farms, farmed fish tend to turn up in salmon streams — in some cases outnumbering their wild counterparts. In western Canada and in Washington state, south of the US border, ecologists are becoming particularly concerned about the effects of escaping Atlantic salmon — which number tens, if not hundreds, of thousands — on already declining Pacific salmon populations.

The salmon run

Whether transgenic salmon pose a special risk is uncertain, but the potential problems are clear enough. One worry is that escaped GM fish will breed with their wild counterparts and release their added growth-hormone genes into wild populations, with unpredictable consequences. Proponents of the technology counter that it is possible to make the transgenic fish sterile — and if Aqua Bounty Farms' Atlantic salmon were farmed off the Pacific coast of British Columbia and the northwest United States, they would be unlikely to breed successfully with native Pacific salmon species.

Even if fast-growing GM fish do not spread their genes to their wild counterparts, they could disrupt the ecology of salmon streams by competing with native fish for resources. The consequences will depend on many factors, including the health of the local population, the number and specific genetic strain of the escaped fish, and the local environment.

On the question of interbreeding, alarming results have come from laboratory studies conducted by William Muir and Richard Howard of Purdue University in West Lafayette, Indiana. Using the fast-breeding Japanese medaka (*Oryzias latipes*) as an experimental model, Muir and Howard looked at the role of size in mating success, and found that big medaka males had a fourfold advantage over their smaller competitors.

The researchers then compared the via-

bility of normal medaka with another group to which they had added a human growth-hormone gene. Under aquarium conditions, the fast-growing GM fish were 30% more likely to die before reaching sexual maturity. The final step was to plug these and other results into a computer to model to see what would happen when 60 transgenic fish were introduced into a population of 60,000 wild medaka. The results were disturbing. It took only 40 generations for the GM fish, which mated more successfully but produced offspring that did not survive as well, to drive the population to extinction. Muir and Howard called it the "Trojan gene effect"².

The Purdue researchers stressed that their results should be treated with caution. Among other things, the dire prediction of population extinction assumed that mature transgenic fish would be bigger than their wild counterparts — whereas the human growth-hormone gene only increased the medaka's juvenile growth rate, and produced adult fish no bigger than average. Muir has since been experimenting with the gene for a salmon growth hormone and has found that it can make adult medaka grow up to 50% larger than normal. The viability of these fish was even worse — their survival to sexual maturity was reduced by as much as 78% compared with wild-type medaka, which suggests that they could wipe out a wild population very quickly. These results have yet to be published, but make the Trojan gene seem like a real threat if the techniques used to make GM fish sterile prove less than 100% reliable.

Shock treatment

Creating sterile salmon is relatively simple. If salmon eggs are subjected to a heat or pressure shock shortly after fertilization, they retain an extra set of chromosomes, ending up with three sets, rather than the normal two. The resulting 'triploid' fish do not

develop normal sexual characteristics, and the females are sterile. It is also possible to raise female salmon as fertile males by treating them with male sex hormones. So by using these 'sex reversed' males — which will be able to produce only female offspring — as breeding stock, and applying a pressure shock to the eggs that they fertilize, it should be possible to raise GM salmon that consist entirely of sterile, triploid females.

Fletcher argues that skilled aquaculturists can apply this method unerringly, but other scientists are less confident. "Even when you're pretty good at it, you get a lot of batch to batch variation," argues Anne Kapuscinski, a specialist in biotechnology and aquaculture at the University of Minnesota in St Paul.

A growing problem?

But even if sterility cannot be guaranteed, will Aqua Bounty Farms' salmon grow into oversized adults that have an advantage in the mating game? That is the "million dollar question", says Kapuscinski. Commercial fish farmers are only interested in having salmon that grow to market size faster, and Fletcher says the company's studies have found "zero evidence" that the transgenic salmon are bigger after they reach sexual maturity. But other scientists point out that these results have not been published in the peer-reviewed literature. "No one outside of their circle has seen those data," complains Eric Hallerman, a fisheries biologist at the Virginia Polytechnic Institute in Blacksburg.

Devlin, who has raised growth-enhanced transgenic coho salmon in the lab, finds that they are about 50% larger at sexual maturity than their wild counterparts. But that may in part reflect the difference between cosy lab conditions and the harsh natural environment. The key test is to grow transgenic and wild-type fish under identical conditions. When scientists at the Center for Genetic Engineering and Biotechnology in Havana, Cuba, conducted such experiments with GM tilapia, the genetically engineered fish grew up to twice as large at maturity as non-transgenic fish³. So far, no one has published this type of experiment with salmon.

Studies addressing the ability of transgenic salmon to disrupt ecosystems irrespective of their ability to interbreed with wild populations have yielded similarly inconclusive results. Devlin and his colleagues, for instance, have found that growth-enhanced transgenic coho salmon eat nearly three times as much food as their natural counterparts under laboratory conditions⁴ — their elevated growth-hormone levels appear to make them hungrier. Whether this would hold true in the wild is uncertain. But if so, the transgenic fish, which also mature faster, could be foraging ravenously at times when natural food availability in a particular stream is low, which could seriously disrupt its ecology.

Arnold Sutterlin of Aqua Bounty Farms, working with Mark Abrahams of the University of Manitoba in Winnipeg, has conducted similar experiments with growth-enhanced Atlantic salmon. Again, the fish were hungrier, but they were also less careful about avoiding predators — judged by experiments in which young fish had to get their food from a portion of a tank containing a large trout⁵. This, together with the observation that young transgenic Atlantic salmon appear to have less effective camouflage, should mean that they are less likely to survive in the wild, minimizing the ecological damage that escaping fish might cause.

Devlin and his colleagues also report that their transgenic coho are slower swimmers⁶. But Aqua Bounty Farms' scientists have found that the GM Atlantic salmon appear more active than the wild-type fish^{5,7}. These varying results may reflect the different species being studied, or just the strain-to-strain variations caused by the vagaries of transgenic technology. Depending on where exactly the extra genes are incorporated into the fish genome, they can exert subtly different effects. For example, although Muir's experiments with a salmon growth-hormone gene in medaka showed that survival to sexual maturity was depressed by up to 78%, in some strains the figure was only 40%. To conduct an accurate environmental risk assessment for GM fish, he says, you need to evaluate each genetic line individually. "We're not sure of all the reasons why," says Muir, "but every transgenic founder is unique."

To Devlin, the catalogue of scientific uncertainties shows why more research is desperately needed. And it will not come cheap. "These are not small experiments," he says. "When you're talking about raising a family of transgenic fish, it's not a vial. It's large tanks." Muir would start by conducting aquarium experiments to test how transgenic fish compare with normal fish over a range of parameters related to their biological 'fitness'. In the manner of his medaka experiments, he would then generate a computer model to consider the likely impact of the fish in wild populations. To this Devlin would add laboratory studies that simulate stream conditions. Even better, says ecologist Jeff Hutchings of Dalhousie University in Halifax, Nova Scotia, would be to conduct tests in the wild, blocking off a portion of a stream to prevent fish escaping into a larger aquatic system. Whether it would be feasible to isolate streams in this way is unclear — and any experiment involving the deliberate environmental release of a transgenic animal is likely to prove highly controversial.

The FDA is still considering Aqua Bounty



Coming ashore: concerns about damage to wild salmon (left) may force fish farms inland (above).

is relatively cheap, costing only 20 cents per fish, claims Kapuscinski. But any regulations preventing GM salmon from being raised in coastal waters would pose problems. Production of Atlantic salmon is presently conducted almost exclusively in sea pens. Unless the performance of transgenic salmon becomes truly remarkable, Fletcher believes it is unlikely that a sizeable portion of the industry would switch to contained land-based, pumped systems given their high capital and operating costs.

But Elliot Entis, chief executive officer of Aqua Bounty Farms, takes a more sanguine view. Even if the FDA imposed a ban on the rearing of growth-enhanced GM salmon in coastal net pens, he believes the transgenic fish could become a viable economic proposition in the long term. The UN Food and Agriculture Organization predicts that global aquacultural production will more than double over the coming decade. Given that coastal aquaculture is already causing ecological damage, by spreading fish diseases, modifying habitats, causing nutrient pollution, and through the escape of exotic farmed fish⁸, Entis believes regulators may eventually demand that fish farms move from coastal pens to contained ponds. "A lot of salmon farming is going to move inland regardless," he says. If so, argues Entis, fast-growing transgenic salmon might be just what the industry needs to remain economically competitive. ■

Tony Reichhardt writes from Washington for Nature.

Farms' application to market its transgenic salmon, and has yet to clarify the experiments it wants to see conducted before deciding whether to approve the fish. The FDA's main task is to examine whether transgenic salmon are fit for human consumption, and whether the fishes' own welfare is compromised by the addition of growth-hormone genes. But the agency is also charged with assessing the environmental impact of the fish, and so it is consulting with two other US government agencies, the Fish and Wildlife Service and the National Marine Fisheries Service.

Given the uncertainties, many ecologists argue that a 'safety first' approach is essential. Kapuscinski would even go so far as to demand that transgenic salmon should be raised in isolated artificial ponds, so they cannot escape to the wild.

Swimming against the tide

For salmon farmers, such measures would impose economic penalties. Because they reach market size faster, Aqua Bounty Farms has touted its GM fish as being superior in terms of 'feed conversion' — how much it costs to feed them while they are being raised. But Fletcher is cautious about putting any firm figures on this. And the advantage would have to be large to counter the costs of the measures suggested by Kapuscinski.

Aqua Bounty Farms is assuming that the FDA will demand it raises only sterile, female triploid fish. Individual screening for sterility

1. Hew, C. L., Fletcher, G. L. & Davies, P. L. *J. Fish Biol. Suppl. A* **47**, 1–19 (1995).
2. Muir, W. M. & Howard, R. D. *Proc. Natl Acad. Sci. USA* **96**, 13853–13856 (1999).
3. de la Fuente, J., Guillen, I., Martinez, R. & Estrada, M. P. *Genet. Anal. Biomol. Eng.* **15**, 85–90 (1999).
4. Devlin, R. H. *et al. Aquaculture Res.* **30**, 479–482 (1999).
5. Abrahams, M. V. & Sutterlin, A. *Anim. Behav.* **58**, 933–942 (1999).
6. Farrell, A. P., Bennet, W. & Devlin, R. H. *Can. J. Zool.* **75**, 335–337 (1997).
7. Stevens, E. D., Sutterlin, A. & Cook, T. J. *Can. J. Fisheries Aquat. Sci.* **55**, 2028–2035 (1998).
8. Naylor, R. L. *et al. Nature* **405**, 1017–1024 (2000).

How can we build a 'knowledge economy' if research is handcuffed?

Sir — New Zealand has found a novel approach to the problems caused by the brain drain and by chronic underfunding for basic biological research^{1,2}. The government's solution is to make much biological research illegal instead of just prohibitively expensive (see page 8).

The Hazardous Substances and New Organisms (HSNO) Act has imposed the most oppressive restrictions on laboratory biological research in the Western world, making even safe science a potentially criminal endeavour. Most universities and research institutions were unaware of the act's full implications until they recently started coming up against them. Just this April, the Environmental Risk Management Authority (ERMA) revoked all previously given institutional authorities to approve laboratory experiments. It halted any new work involving genetic modification, even the most trivial, until ERMA itself had inspected each experiment.

HSNO makes no distinction between, on the one hand, routine genetic manipulation such as making a DNA library, performed in containment laboratories designed to keep organisms in and the public out, and, on the other, genetic manipulation designed to produce a biological product or organism for field release. Every procedure that puts DNA with any history of *in vitro* manipulation into any 'organism' has to go through an expensive risk assessment by ERMA. The assessment must anticipate every manipulated nucleic acid and every recombinant organism: imagine that, if you are making a DNA library.

ERMA requires up to NZ\$3,000 (US\$1,400) for each application to import modified organisms safely into containment, even if they are only a source of a new cloning plasmid. So research organisms identical to those that might be considered safe if developed in New Zealand — where there is not enough money to fund their development — are too expensive to import. Materials developed here will still be distributed freely as part of the generosity shown by scientists to one another, but we will not be able to afford to import the analogous materials offered to us for free.

As a university lecturer, I think it would be irresponsible of me to not advise New Zealand biology students to leave for an overseas education. In the short term, we cannot afford to teach them modern genetic techniques — not because our laboratories are unsafe, but because of the ongoing compliance costs under HSNO. In the long term, we will neither keep those most qualified

to teach modern biology, nor attract many of our best young biologists back.

Ironically, our government is advocating a new 'knowledge economy'. How this is to be created is not clear.

Jack A. Heinemann

Department of Plant and Microbial Sciences,
University of Canterbury, Private Bag 4800,
Christchurch, New Zealand

1. Heinemann, J. A. *Drug Discovery Today* **5**, 222–223 (2000).
2. Swinbanks, D. *Nature* **379**, 112 (1996).

Climate-change analysis has been changing too

Sir — The valuable Commentary by Hans von Storch and Nico Stehr¹ is of critical importance for what it tells us about the cultural context of climate change and impacts research. However, we do not agree with some of the authors' more rhetorical statements.

First, the authors go from a description of how climate change has been perceived in the past to a claim that current impact assessments are simplistic: a scientific judgement. Surely the historical facts don't tell us whether or not a contemporary climate impact analysis is simplistic.

Second, the authors state that climate change is occurring, but that claims about significant future impacts are only a hypothesis. Of course, the science of analysing climate-change impacts is in its early stages, and there is much more uncertainty about the precise nature of those impacts than about whether climate change will happen. But this is not because our information about future impacts is hypothetical whereas that about current climate changes is not. There is a vast literature in the climate-impacts field that criticizes past work as simplistic and argues for more sophisticated approaches (for example, discussions about the 'naive farmer' hypothesis). This is a normal progression, not a sign of fatal weakness.

Studies on global water and food security^{2,3} and regional case studies on impacts and adaptation^{4–8} illustrate that the dialogue has become much more mature. The 'apocalyptic scenario' has been replaced by genuine concerns about how climate change will exacerbate food, water and health problems in many countries.

The real value of the historical analysis in the Commentary is the cultural context it provides for research on climate change and impacts. In turn, this gives us an important indication of how the climate issue may play out in the public arena; about how it may be interpreted by politicians, members of the public and others; and about possible reactions and responses to climate-policy initiatives. To grapple effectively with these questions, we need

more integrated analysis that takes into account such historical, philosophical and social-scientific questions.

John B. Robinson and Stewart J. Cohen*

Sustainable Development Research Institute,
University of British Columbia, 2202 Main Mall,
Vancouver V6T 1Z4, Canada

* also at the Adaptation & Impacts Research Group,
Meteorological Service of Canada, Environment
Canada, 2029 West Mall, Vancouver V6T 1Z2,
Canada

1. von Storch, H. & Stehr, N. *Nature* **405**, 615 (2000).
2. Arnell, N. W. *Global Environmental Change* **9**, S31–S51 (1999).
3. Downing, T. E. (ed.) *Climate Change and World Food Security* (Springer, Heidelberg, 1996).
4. Cohen, S. J. *Arctic* **50**, 293–307 (1997).
5. Yates, D. N. & Strzepek, K. M. *Climatic Change* **38**, 261–287 (1998).
6. Nishioka, S. & Harasawa, H. *Global Warming: The Potential Impact on Japan* (Springer, Tokyo, 1998).
7. University of Washington. *Impacts of Climate Variability and Change in the Pacific Northwest* (JISAO/SMA Climate Impacts Group, Seattle, 1999).
8. Huq, S., Karim, Z., Asaduzzamana, M. & Mahtab, F. (eds) *Vulnerability and Adaptation to Climate Change in Bangladesh* (Kluwer, Dordrecht, 1999).

Unrealistic promises on AIDS fuel scepticism

Sir — Incredulous as I am about South African President Thabo Mbeki's ill-informed perspective on HIV and AIDS, I am even more incredulous that, according to your News report "Letter fuels South Africa's AIDS furor" (*Nature* **404**, 911; 2000), the president of the World Bank, James Wolfensohn, "promised that there would be 'no limit' to the funds available for combating AIDS in the developing world".

Does this mean the World Bank intends to assume the entire cost of AIDS medicine and medical care for all those in the developing world? I very much doubt it. It is precisely this kind of claim — reported without qualification — that makes Mbeki and many others regard Western pronouncements on AIDS as untrustworthy.

Your reporter's description of the right-wing Boerestaat party's support for Mbeki's stance as coming "from an unusual quarter" overlooks the well-reported right-wing sentiments of several prominent US opponents of the HIV–AIDS connection. Many right-wingers see AIDS as a "moral" disease, caused by degenerate sexual and drug habits. It is not surprising that those sharing this political perspective have opposed efforts, such as clean-needle programmes, aimed at resisting the spread of this scourge.

To minds such as these, death from AIDS is the appropriate remedy for immoral behaviour. Those who support the South African right wing are also unlikely to feel distressed at the prospect of an AIDS-depopulated black Africa.

Dan Dorritie

PO Box 738, Davis, California 95617, USA

The declaration on these two pages was stimulated by the current controversy in South Africa about whether HIV is the cause of AIDS (see, for example, *Nature* 404, 911 & 405, 105; 2000). This has caused massive consternation among all scientists, doctors and many others in the international community who treat AIDS patients or who work on AIDS in other ways. There is widespread anxiety that denying or doubting the cause of AIDS will cost countless lives if blood screening, use of condoms, and methods to prevent mother-to-child transmission of the virus are not implemented or, worse, even abandoned.

The declaration has been signed by over 5,000 people, including Nobel prizewinners, directors of leading research institutions, scientific academies and medical societies, notably the US National Academy of Sciences, the US Institute

of Medicine, Max Planck institutes, the European Molecular Biology Organization, the Pasteur Institute in Paris, the Royal Society of London, the AIDS Society of India and the National Institute of Virology in South Africa. In addition, thousands of individual scientists and doctors have signed, including many from the countries bearing the greatest burden of the epidemic. Signatories are of MD, PhD level or equivalent, although scientists working for commercial companies were asked not to sign.

The Durban Declaration has an organizing committee of over 250 members from over 50 countries. The list of signatories up to 29 June can be found on *Nature's* website as Supplementary Information (<http://www.nature.com>), and an up-to-date list can be found at <http://www.durbandeclaration.org>.

The Durban Declaration

HIV causes AIDS. Curbing the spread of this virus must remain the first step towards eliminating this devastating disease.

Seventeen years after the discovery of the human immunodeficiency virus (HIV), thousands of individuals from around the world are gathering in Durban, South Africa, to attend the XIII International AIDS Conference, which starts next week (9 July). At the turn of the millennium, figures released last week reveal that an estimated 34.3 million people worldwide are living with HIV or AIDS, 24.5 million of them in sub-Saharan Africa¹. Last year alone, 2.8 million people died of AIDS, the highest rate since the start of the epidemic. If current trends continue, southern and Southeast Asia, South America and regions of the former Soviet Union will also bear a heavy burden in the next two decades.

AIDS spreads by infection, like many other diseases, such as tuberculosis and malaria, that cause illness and death particularly in underprivileged and impoverished communities. HIV-1, which is responsible for the AIDS pandemic, is a retrovirus closely related to a simian immunodeficiency virus (SIV) that infects chimpanzees. HIV-2, which is prevalent in West Africa and has spread to Europe and India, is almost indistinguishable from an SIV that infects sooty mangabey monkeys. Although HIV-1 and HIV-2 first arose as zoonoses² — infections transmitted from animals to humans — both now spread among humans through sexual contact; from mother to infant; and via contaminated blood.

An animal source for an infection is not unique to HIV. The plague came from rodents and influenza from birds. The new Nipah virus in Southeast Asia reached humans via pigs. Variant Creutzfeldt-Jakob disease in the United Kingdom is identical to 'mad cow' disease. Once HIV became established in humans, it soon followed human habits and movements. Like many other

viruses, HIV recognizes no social, political or geographic boundaries.

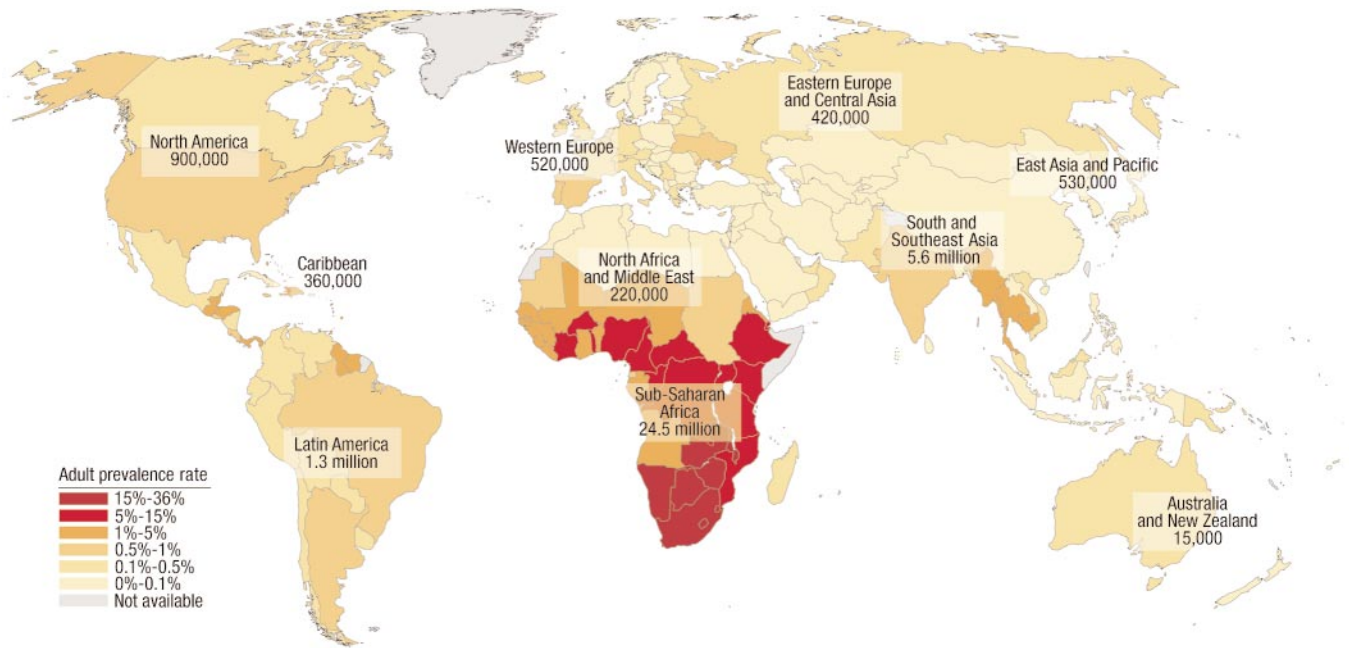
The evidence that AIDS is caused by HIV-1 or HIV-2 is clear-cut, exhaustive and unambiguous, meeting the highest standards of science³⁻⁷. The data fulfil exactly the

same criteria as for other viral diseases, such as polio, measles and smallpox:

- Patients with acquired immune deficiency syndrome, regardless of where they live, are infected with HIV³⁻⁷.



Future orphans? The death toll from AIDS in Africa will be enormous unless action is taken now.



Adults and children estimated to be living with HIV/AIDS, together with the proportion of adults infected by HIV across the world (UNAIDS, June 2000).

● If not treated, most people with HIV infection show signs of AIDS within 5–10 years^{6,7}. HIV infection is identified in blood by detecting antibodies, gene sequences or viral isolation. These tests are as reliable as any used for detecting other virus infections.

● People who receive HIV-contaminated blood or blood products develop AIDS, whereas those who receive untainted or screened blood do not⁶.

● Most children who develop AIDS are born to HIV-infected mothers. The higher the viral load in the mother, the greater the risk of the child becoming infected⁸.

● In the laboratory, HIV infects the exact type of white blood cell (CD4 lymphocytes) that becomes depleted in people with AIDS^{3–5}.

● Drugs that block HIV replication in the test tube also reduce virus load in people and delay progression to AIDS. Where available, treatment has reduced AIDS mortality by more than 80% (ref. 9).

● Monkeys inoculated with cloned SIV DNA become infected and develop AIDS¹⁰.

Further compelling data are available⁴. HIV causes AIDS⁵. It is unfortunate that a few vocal people continue to deny the evidence. This position will cost countless lives.

In different regions of the world, HIV/AIDS can show altered patterns of spread and symptoms. In Africa, for example, people infected with HIV are 11 times more likely to die within five years⁷, and

more than 100 times more likely than uninfected people to develop Kaposi's sarcoma, a cancer linked to yet another virus¹¹.

As with any other chronic infection, various factors have a role in determining the risk of disease. People who are malnourished, who already suffer other infections or who are older, tend to be more susceptible to the rapid development of AIDS following HIV infection. However, none of these factors weakens the scientific evidence that HIV is the sole cause of the AIDS epidemic.

In this global emergency, prevention of HIV infection must be our greatest worldwide public-health priority. The knowledge and tools to prevent infection are available. The sexual spread of HIV can be stopped by mutual monogamy, abstinence or by using condoms. Blood transmission can be prevented by screening blood products and by not reusing needles. Mother-to-child transmission can be reduced by half or more by short courses of antiretroviral drugs^{12,13}.

Limited resources and the crushing burden of poverty in many parts of the world constitute formidable challenges to the control of HIV infection. People already infected can be helped by treatment with life-saving drugs, but the high cost of these drugs puts these treatments out of reach for most of the world. It is crucial to develop new antiviral drugs that are easier to take, have fewer side effects and are much less expensive, so that millions more can benefit from them.

There are many ways of communicating the vital information on HIV/AIDS, and what works best in one country may not be appropriate in another. But to tackle the disease, everyone must first understand that HIV is the enemy. Research, not myths, will lead to the development of more effective

and cheaper treatments, and, it is hoped, a vaccine. But for now, emphasis must be placed on preventing sexual transmission.

There is no end in sight to the AIDS pandemic. But, by working together, we have the power to reverse its tide. Science will one day triumph over AIDS, just as it did over smallpox. Curbing the spread of HIV will be the first step. Until then, reason, solidarity, political will and courage must be our partners.

1. Joint United Nations Programme on HIV/AIDS (UNAIDS) *Report on the Global HIV/AIDS Epidemic, June 2000* (UNAIDS, Geneva, 2000). <http://www.unaids.org/hiv/aidsinfo/documents.html>
2. Hahn, B. H., Shaw, G. M., De Cock, K. M. & Sharp, P. M. AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607–614 (2000).
3. Weiss, R. A. & Jaffe, H. W. Duesberg, HIV and AIDS. *Nature* **345**, 659–660 (1990).
4. NIAID *HIV as the Cause of AIDS* <http://www.niaid.nih.gov/spotlight/hiv00/>
5. O'Brien, S. J. & Goedert, J. J. HIV causes AIDS: Koch's postulates fulfilled. *Curr. Opin. Immunol.* **8**, 613–618 (1996).
6. Darby, S. C. *et al.* Mortality before and after HIV infection in the complete UK population of haemophiliacs. *Nature* **377**, 79–82 (1995).
7. Nunn, A. J. *et al.* Mortality associated with HIV-1 infection over five years in a rural Ugandan population: cohort study. *Br. Med. J.* **315**, 767–771 (1997).
8. Sperling, R. S. *et al.* Maternal viral load, zidovudine treatment, and the risk of transmission of human immunodeficiency virus type 1 from mother to infant. *N. Engl. J. Med.* **335**, 1678–1680 (1996).
9. Centers for Disease Control and Prevention (CDC) *HIV/AIDS Surveillance Report 1999* **11**, 1–44 (1999).
10. Liska, V. *et al.* Viremia and AIDS in rhesus macaques after intramuscular inoculation of plasmid DNA encoding full-length SIVmac239. *AIDS Res. Hum. Retroviruses* **15**, 445–450 (1999).
11. Sitas, F. *et al.* Antibodies against human herpesvirus 8 in black South African patients with cancer. *N. Engl. J. Med.* **340**, 1863–1871 (1999).
12. Shaffer, N. *et al.* Short course zidovudine for perinatal HIV-1 transmission in Bangkok Thailand: a randomised controlled trial. *Lancet* **353**, 773–780 (1999).
13. Guay, L. A. *et al.* Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial. *Lancet* **354**, 795–802 (1999).

A call for a chlorine sunset

The hazards of organochlorines require that they be eliminated.

Pandora's Poison: On Chlorine, Health, and a New Environmental Strategy

by Joe Thornton

MIT Press: 2000. 611 pp. \$34.95, £23.50

Terry Collins

Organochlorines are at the centre of a fierce, high-stakes debate. This contribution to the debate establishes Joe Thornton, a scholar at Columbia University and former Greenpeace scientist, as the standard-bearer for those who consider that many anthropogenic organochlorines are a source of environmental and health hazards.

Thornton argues, in his meticulously researched, 11-chapter book, that a "chlorine sunset" must be orchestrated. He contends that the products and processes of the chlorine industry seriously compromise the environment as a safe place for maintaining the fertility and abundance of life.

At the opposite pole in the debate stand the Chlorine Chemistry Council (C-3), a subgroup of the Chemical Manufacturers Association, and the Vinyl Institute (VI), a consortium of the makers and users of vinyl. The C-3 position has been presented recently by Clifford T. Howlett Jr, who argues in *Chlorine and Chlorine Compounds in the Paper Industry* (Ann Arbor Press, 1997) that chlorine is vital for health and the economy; Howlett ridicules the Thorntons of this world and their case.

Ironically, Howlett's tactics stimulated the writing of *Pandora's Poison*, the latest and by far the most potent salvo in the conflict. Thornton sets out to refute the C-3 charge that the industry's opponents don't practise 'sound science'. He points out that this accusation is, in fact, the fulcrum of a public relations campaign to dismiss the industry's critics by implying that they base their judgments on "bad science or — even worse — emotion, fear or some other suspect motive".

As the centre-piece to his case, Thornton criticizes the current approach to the regulation of chemicals, calling it the "risk paradigm". According to this, a chemical is regulated or banned only when science establishes that the chemical presents what a regulatory agency deems to be an unacceptable risk. Thornton argues that this is "utterly ill-suited to addressing the long-term global health threat that organochlorines pose". He proposes instead a new regulatory dynamic that he calls the "ecological paradigm". This would be based on the "precautionary principle ... [that] ... we should avoid practices that have the potential to cause severe harm, even in the absence of scientific proof of harm". The



The labyrinthine infrastructure of the chlorine economies.

precautionary principle would be exercised under guidance principles called "zero discharge", "clean production" and "reverse onus". Chlorine economies in water disinfection and the enormous, labyrinthine markets for vinyl, chemicals, paper, surface cleaning and clothes bleaching are thus pitted against the negative impact that organochlorines may have on the health and welfare of all life-forms.

Pandora's Poison is a landmark book which should be read by anyone wanting to understand the environmental and health dangers of the chlorine industry. As a reference work alone it is a masterpiece, analysing around 1,000 references. In areas with which I am familiar, Thornton's treatment is mostly brilliant and is based on sound science.

But it is much more than this. Thornton develops his case comprehensively, evaluating the landscape from the historical, through the environmental and scientific, to the industrial, political, regulatory and ethical forces that are now central to the chlorine struggle. He argues convincingly that widely distributed, persistent organochlorines are a major global hazard, and he backs his arguments up with an eloquent explanation of the chemical properties that determine the dangers. The cases for and against many organochlorines as potent toxins and carcinogens are detailed. Thornton discusses studies which reveal how certain organochlorines can significantly disrupt the endocrine system, acting in vanishingly small quantities. Such disruption can compromise life and fertility in largely invisible ways that can be extraordinarily dangerous. And there are proven cases of organo-

chlorines that have severely harmed both humans and animals, and strong connections of some organochlorines with cancer.

There have been few studies of the toxicology of organochlorines or, more importantly, of the toxicology of combinations of organochlorines. Thornton is strongly critical of toxicologists as well as of the chemical industry over this, as he believes the problem to be much greater than is currently perceived. We now know enough about the unintended perils of persistent compounds to understand that this argument is unimpeachable.

Thornton explains the history behind the growth of the industrial use of chlorine over the past 60 years, describing how it has come to represent a huge part of the economy. Thornton cites shocking evidence of major chemical companies distorting scientific data so as to make it appear that dioxins are harmless to humans. Such action betrays the wider chemical community, discrediting all chemists through the unscrupulous behaviour of people who represent chemistry. Moreover, these falsified studies appear to be heavily responsible for the widespread underestimation of the dangers of dioxins to humans.

Similarly, Thornton discloses a Vinyl Institute memorandum laying out plans to stave off the implementation of punitive regulations being planned by the Environmental Protection Agency for dioxin production associated with the incineration of vinyl. VI leadership decided to hide behind the prestige of the American Society of Mechanical Engineers. It funded a consultant who was "willing to set his priorities to [VI] needs", producing the answers they wanted disguised as an

ASME report. If Thornton has been given the complete story accurately, the whole dishonourable affair would undermine the credibility of this industrial group as a body interested in honest scientific dialogue about the effects of their products on human health.

In *Pandora's Poison* Thornton is magnificent, but he is not perfect. He tends to give chlorine no quarter, sometimes polarizing the arguments in the book more than is necessary or desirable (although balance is usually achieved). Some of his suggestions for alternative technologies can be easily criticized, and I don't always agree with his assessments of how the industrial landscape can be changed to eliminate organochlorines.

Readers must judge for themselves how we can best attain a sustainable future. But, in my opinion, it is definitely time to thank the chlorine industry for useful but imperfect technologies developed when the hidden dangers were not understood, and prepare to move on expeditiously, in particular by developing safer alternatives and exploring Thornton's policy suggestions.

Pandora's Poison also has important indirect lessons for achieving a sustainable civilization. If young chemists were rigorously taught about the toxicity of the substances they synthesize and study, their efforts would more naturally be directed towards avoiding persistent toxic substances. The chemical vernacular would include the terms around which *Pandora's Poison* is constructed — for example, persistent toxins (both elemental and molecular), endocrine disruption, bioaccumulation, atmospheric distillation, mutagen and teratogen.

One of the most pressing tasks for a sustainable civilization is to replace technologies producing persistent pollutants with safe alternatives. As Thornton says of our use of anthropogenic organochlorines, "like the Romans, who sipped from lead cups, ran drinking water through lead pipes, and bathed in lead basins, we have built our house of poison unaware of the consequences". ■

Terry Collins is in the Department of Chemistry, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213, USA.

Sexual trade-offs

Promiscuity: An Evolutionary History of Sperm Competition and Sexual Conflict

by Tim Birkhead

Faber & Faber: 2000. 272 pp. £9.99 (pbk)

A. H. Harcourt

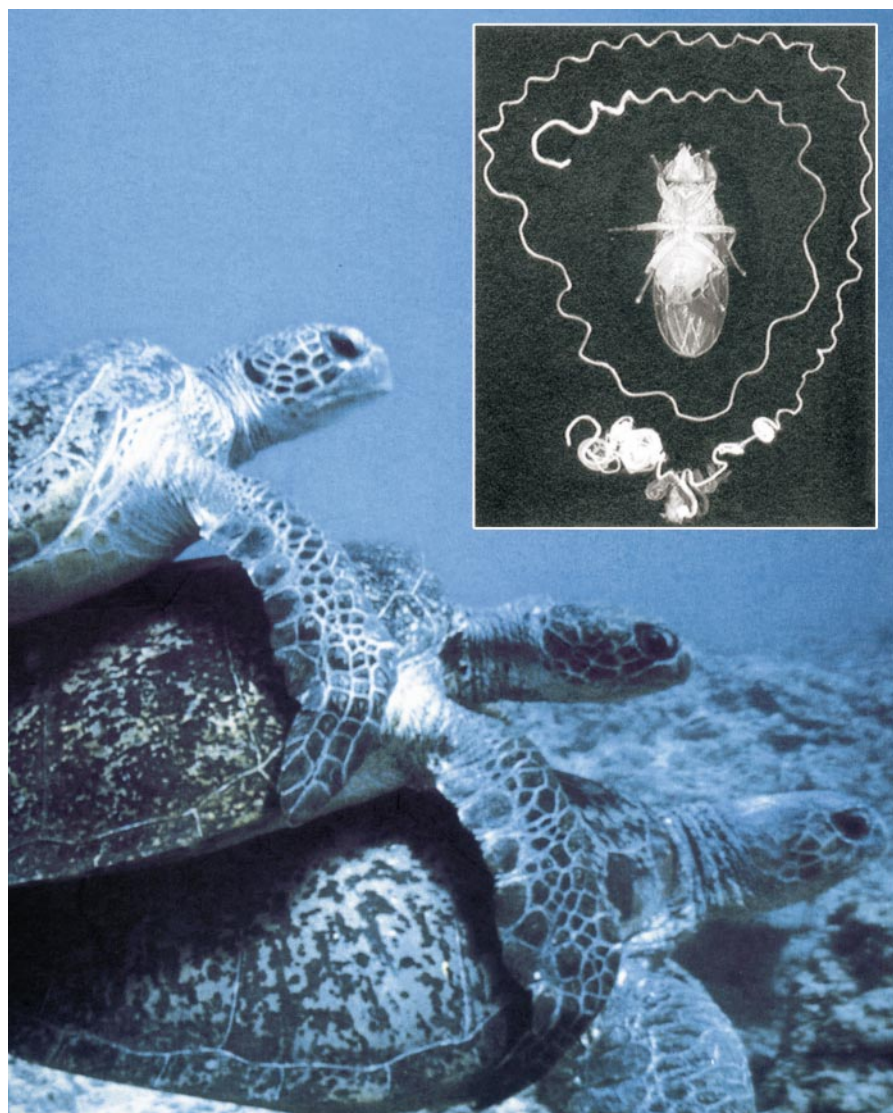
In some barnacle species, males are mere bags of sperm in special pouches inside the female. Fourteen bags (males) in one female was the highest number Darwin counted. He didn't expand on such promiscuity in females, perhaps because his prim daughter Henrietta was by then proofreading and editing his manuscripts. But Tim Birkhead more than makes up for Darwin's omission. *Promiscuity* is a fascinating, wide-ranging, erudite, readable journey through some of the weirder stretches of biology.

Promiscuous behaviour has some amazing manifestations. Many of them, for males, are well known. Deers' antlers and peacocks' tails are obvious examples; others we are less familiar with. Perhaps most surprising is the extent of promiscuity in females. Birkhead tells us of a feral female sheep that mated with seven males in five hours, for 163 matings in total. But the evidence for female promiscuity has been before our eyes for decades. Witness the huge, pink sexual swellings on the rumps of the females of many primate species. If such swellings were on males, they would surely have received extensive attention in Darwin's *The Descent of Man, and Selection in Relation to Sex*. As it is, we had to wait until the 1970s for their association with female promiscuity to be highlighted and explained in terms of the advantages for females in attracting several mates.

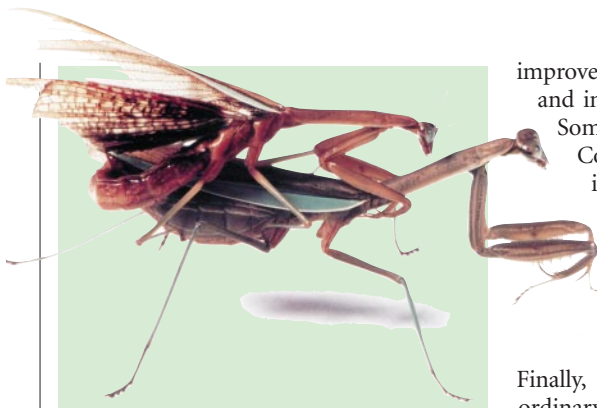
We are still uncertain as to what these advantages are. Author and anthropologist Sarah Hrdy's long-standing suggestion for promiscuity by female primates is that it reduces the number of males that might mistreat the female's offspring. So, in one human society, in which infants have a 50% chance of dying if their father dies or departs, the offspring of mothers who have more than one male accepting fatherhood survive better than do the offspring of monogamous mothers, because the infants are not perceived as fatherless and hence defenceless.

Testing of this and other hypotheses is a rapidly growing area of some very nice research. Thus, promiscuous female pseudoscorpions abort fewer eggs than do monogamous females, because the promiscuous ones are more likely to have mated with a genetically compatible partner.

Despite the female monkey's red rump, the main consequences of promiscuity are seen in the male, rather than the female. Whereas female promiscuity necessarily leads to competition among sperm to fertilize



Turning turtle, and (inset) the longest sperm, from the fruitfly.



Mate now, meal later

Sex with a female preying mantis can end in death. From *Food Chain: Encounters Between Mates, Predators, and Prey* (Aperture, \$29.95, £20), photographs by Catherine Chalmers.

the egg, the reverse is not so clearly the case (although males can run out of sperm). Therefore, if a female mates with more than one male, it will be to each male's advantage to inseminate more sperm than if the female were monogamous.

The reason is simple. Fertilization is often a lottery, and the males with the most tickets (sperm) are most likely to win. Consequently, in species in which females are especially promiscuous, males have particularly large testes, which produce extremely high numbers of good-quality sperm: males of the polyandrous chimpanzee have testes that are 16 times the size of those of the far less promiscuous gorilla.

And the story gets a lot more bizarre. The copulating male redback spider places itself on the female's jaws, and hence is eaten, apparently because, by becoming a meal, he prolongs copulation, extends the female's delay in remating, and hence diminishes the chances of a rival mating with her.

Chemicals in fruitfly semen, some similar to spider toxin, induce egg production in females, and delay subsequent mating with another male. Probably as an unfortunate side effect, they also reduce her lifespan. But as long as she has bred by then, and she will have done, her shorter life is of no cost to the male. And in a lovely experiment to show that the toxicity of the fruitfly's semen is a competitive consequence of the female's promiscuity, when females and males are bred monogamously for a few generations, the semen ceases to be toxic.

It is a fruitfly that holds the record for the longest sperm known. The fly itself (*Drosophila bifurca*) is just 1.5 millimetres long, a normal size for a fruitfly. But its sperm is six centimetres long — yes, centimetres. The picture of the male fly surrounded by several coils of its sperm is one of the more memorable among a number of startling pictures in the book.

Females have evolved adaptations both to

improve their chances of being promiscuous and in response to adaptations in males.

Some promiscuity is forced on females.

Competing bedbug males go as far as to inject sperm through the female's abdominal wall in attempts to beat other males to her ova. Her response in the battle to gain control is, it seems, to develop pads of tissue under the cuticle that look as if they absorb and kill the sperm.

Finally, take the rove beetle, an otherwise ordinary insect. The male inseminates the female with a bag of sperm which, once inside her body, extrudes a tube into her sperm store. The end of the tube expands into a bulb, forcing out sperm from previous males. The bulb is then pierced by two 'teeth' that extrude from the female's store, releasing the new male's sperm, a lovely story of co-adaptation between the sexes.

I have only two regrets about the book. One is that the index is so sparse, as seems to be the case for too many books nowadays. The other is that, although human promiscuity and its consequences are discussed, 'human' does not appear in the index, not even in the special index of animal species mentioned in the text. The omission is a pity. The book not only corrects some misapprehensions about human promiscuity, but is engagingly enough written for many readers initially interested only in humans to carry on reading. Whole-animal biology needs more appreciation than it receives, and where better a place to start than in an expertly guided walk on the wilder side of reproductive adaptations?

A. H. Harcourt is in the Department of Anthropology, University of California at Davis, Davis, California 95616, USA.

Misplaced nostalgia for a bygone era?

Between Politics and Science: Assuring the Productivity and Integrity of Research

by David H. Guston

Cambridge University Press: 2000. 213 pp. £35, \$54.95

Daniel S. Greenberg

Science in the United States has its own legend of paradise lost, a descent from a halcyon era of political faith in the honesty and productivity of science to a latter-day regime of distrustful oversight and utilitarian demands.

David Guston, assistant professor of public policy at Rutgers University in New Jersey, has produced a skilfully argued and provocative formulation of the American experience. He postulates that, starting around 1980, the

federal government's relationship with science has shifted from benign and distant delegation of authority to assertive micro-management. Central to this change, he contends, was the erosion of a tacit 'social contract' that had congenially bound the parties since the Second World War. In the words of an unnamed government official, the contract provided that "the government will give scientists money to do what they want to do; in return, scientists will try to work on things that are going to be good for ... the people whose money they're spending".

However, "by the early 1980s", says Guston, "the United States decided it could no longer rely entirely on the scientific community to have integrity and be productive all on its own. The political patrons found flaws in the premises of the social contract for science."

Guston attributes the change to doubts that science was serving the nation's economic needs, and to public revelations of scientific misconduct. The government responded with measures to ensure scientific integrity and to encourage technology transfer from government-financed basic research.

Focusing on what he describes as a representative segment of the science-government relationship, Guston observes that the National Institutes of Health established an Office of Scientific Integrity and an Office of Technology Transfer. Congress reserved a portion of science budgets for the programme of Small Business Innovation Research. Other measures provided money and incentives for researchers to think commercially.

Precursors of change, according to the author, occurred in the early 1970s, when Congress and the Nixon White House actively supported applied research at the National Science Foundation. The big difference in the 1980s, Guston asserts, was a shift from macro- to micromanagement through the creation of formal mechanisms at the boundary between science and government, to ensure integrity and technology transfer. The new environment even "empowers whistle-blowers who can publicize information about the failures of scientific integrity". The government still trusts the scientific community, Guston acknowledges, but the "boundary organizations" embody the principle of "trust but verify".

The tale as told is accurate, and the analysis of the shifting relationship between patron and beneficiary persuasive. But, as a science-policy journalist who has continuously observed science-government relations long before and long after the changes at the boundary two decades ago, I find Guston imposes too much order on an amorphous landscape, that he confuses congressional bombast and posturing with tangible effects, and uncritically accepts the mythology of a long-ago Eden in government dealings with science.

In the context of an American economy that spends more than \$225 billion a year on research and development, the few under-

book reviews

staffed, under-financed government outposts set up to promote productivity and integrity play scant roles in scientific affairs. Government itself has receded to a junior partnership in the research economy, providing about 25% of national R&D expenditures.

The misconduct follies played out on Capitol Hill in the 1980s — the so-called Baltimore case and the contention over Robert Gallo's role in identifying the AIDS virus — were the pet project of one powerful House member, John Dingell, and did not come from a groundswell of political concern. Few other legislators seemed interested in scientific misconduct. After the Republicans took control of the Congress in 1995, misconduct disappeared from the legislative agenda. It has reappeared with the recent outbreak of concern over ethical corner-cutting in gene-therapy trials. The Office of Research Integrity in the Department of Health and Human Services has been headed by an acting director since 1996 — scarcely a sign of political approbation for its work. Whistle-blowing is not conducive to career advancement, empowered or not.

As for the social contract, it can just as easily be argued that it never existed, or that, if it did, it persists to this day relatively intact, perhaps even expanded from its original terms. Ironically, these are particularly congenial and trustful times in science-government relations compared with the presumed nirvana of the pre-1980s. Congress today is gung-ho over a rapid doubling of the budget for the National Institutes of Health, based largely on naive faith in that agency's productivity.

Guston's use of language falls victim to academic opacity: "the dissemination model of the second period of Devine, *et al.* remains predicated on the univariate nature of the appropriability model of the first period." Nonetheless, a dogged reading yields fresh insights into the complexities of the American experience in the relations between science and government. ■

Daniel S. Greenberg is at 3736 Kanawha Street, NW Washington DC 20015, USA.

Turning the key to an adolescent talent

In Code: A Mathematical Journey

by Sarah Flannery, with David Flannery
Profile Books: 2000. 292 pp. £14.99

John L. Casti

Sarah Flannery is a teenager in County Cork, Ireland. She is also the creator of a coding scheme for information transmission that dramatically extends our ideas of how best to compress information. How could a completely typical Irish teenager astound the world of mathematics in this way? *In Code* is a first-hand account of the answer.

Sarah's father, mathematics teacher David Flannery, plays the Henry Higgins to Sarah's Eliza Doolittle. The Flannery household, rather ordinary by most standards, is quite unusual in one way: the presence of a large blackboard in the kitchen, on which Flannery senior would write challenging mathematical puzzles for his children to ponder. So, from the time she was a toddler, Sarah was continually exposed to logical thought processes and the thrill of discovery. That's one piece of the answer to her success, a mathematically friendly home environment where the fact that she is a child and a woman played no role in discouraging her from developing a native talent for mathematical thinking.

As Sarah's story continues, we find her taking educational enrichment courses taught by her father at the local college. By the account given in the book, these courses are rather extraordinary in the way they challenge the students to think through the logic behind various mathematical problems. Thus, the principles involved in finding answers are arrived at in a kind of Socratic process of dialogue and discovery.

During the development of her novel coding scheme, Sarah takes the reader through a sequence of science competitions, first in Ireland, then abroad, at each stage of which her project wins a major prize. The ultimate project is a code that improves upon the standard RSA coding scheme used around the world to compress and send information. Sarah's scheme, the technical details of which are not presented in this book, provides an alternative that is considerably faster than the RSA procedure, and thus has the potential to send information far more efficiently — and cheaply — than previously thought possible.

The success of Sarah's coding method in science fairs around the world brings her fame, minor fortune and lots of publicity — including a front-page write-up in *The Times* of London. It also brings innumerable offers from entrepreneurs, software houses and others of that ilk, who promise riches beyond her wildest imagination if she will enter into a commercial arrangement with them to further develop and market her work. To her credit, Sarah refuses all these blandishments, and tells the world that she intends to present her work publicly, essentially giving the code away for free.

But there is a fly in the ointment. As mathematicians scrutinize Sarah's work, they discover a security flaw in the scheme. While the flaw in no way invalidates the mathematical basis of the code, it does prevent it from being used as a public-key cryptosystem. This, in turn, destroys the code's commercial value. *Sic gloria transit mundi*. Sarah remains undaunted by this development, and we read of her admirable aplomb in shrugging off the tarnishing of her achievement by this



Just follow the instructions

Open Here: *The Art of Instructional Design* (Thames & Hudson/Stewart Tabori & Chang, £17.95/\$29.95) by Paul Mijksenaar and Piet Westendorp contains an assortment of visual instructions designed, with varying degrees of success, to help us get through the obstacle course that is everyday life.

blemish. 'Who cares?' she seems to say. Much mathematical work is less than perfect. But it is still regarded as a contribution to progress.

In Code is a wonderfully moving story. While at times it reads a bit too much like a gushy teenager's diary (which it is), the book contains a wealth of interesting information on mathematical puzzles, coding methods, elementary number theory and algebra. It is also well worth noting, however, that it can be profitably read by anyone; no knowledge of mathematics, codes, number theory or anything else is needed. In fact, the book does an exemplary job of walking the reader through a set of graded puzzles aimed at developing mathematical intuition, followed by a first-rate, gentle introduction to codes, deciphering and cryptosystems. So don't be put off by the fear that this is a book on some mathematical genius that you won't understand. It is just the opposite; it's a book about the thrill of the mathematical chase, and how it is a game that anyone can play.

The book also gives a fascinating account of how a gifted teacher like Sarah's father, David, can help nurture and develop the mind of a very bright — but far from genius-level — teenager such as his daughter, and inspire that mind to creative heights one would believe possible only of bona fide geniuses. Sarah's story should serve as an inspiration to all young people, especially young women, who might be contemplating a life in mathematics. I recommend it highly as summer reading not only for teenagers, but for anyone interested in the human spirit and its boundless capacity for innovation and imagination. ■

John L. Casti is at the Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA; also at IASA, A-2361 Laxenburg, Austria, and the Austrian National Research Center, A-2444 Seibersdorf, Austria.

Epistemology engines

An antique optical device has powered several centuries of scientific thought.

Don Ihde

The increasing popularity of the term 'technoscience' as a description of the relations between science and technology is also suggestive of other ways in which science and technology are entwined. Historians of science have a saying: "Science owes more to the steam engine than the steam engine owes to science." Historically, the steam engine developed without much explicit use of scientific theory; yet it inspired the ideas of entropy and the second law of thermodynamics. The machine, not raw nature, suggested the phenomena.

The steam engine shows how a technology can serve as a partial 'epistemology engine'. But a nearly forgotten optical device from early modern science gives a much fuller model for how knowledge itself is produced, a true epistemology engine. This is the camera obscura, which later evolved into the pinhole camera. The optical effect in which some external scene, under light, could be seen as an inverted image inside a darkened room on a blank screen, may have been known to Euclid, but it was clearly described by the Islamic philosopher Alhazen, in his *Optics* of 1037.

Camera obscuras, and related camera lucidas, became well known in the Renaissance: in about 1430, Leon Battista Alberti used them to trace objects with astonishing verisimilitude; the camera contributed to the development of perspective drawing. The optical effect also automatically reduced three dimensions to two. Leonardo da Vinci again described the darkroom (in about 1450) and explicitly made the camera a model for the eye: "when the images of illuminated bodies pass through a small hole into a dark room ... you will see on the paper all those bodies in their natural shapes and colors, but they will appear upside down and smaller ... the same happens inside the pupil."

But the camera did not become a full epistemology engine until both René Descartes and John Locke explicitly made it thus in the seventeenth century. Descartes in *La Dioptrique* and Locke in the *Essay on Human Understanding* both draw upon the camera obscura as a model for how knowledge is produced. For them it is more than the eye that represents the world; the camera is to the eye as the eye is to the mind. Locke's analogues are strikingly literal: "external and internal sensation are the only passages I can find of knowledge to the understanding. These are ... the windows by which light is let



For Descartes and Locke, the camera obscura is more than the eye that represents the world; it is to the eye as the eye is to the mind.

into this dark room: for methinks the understanding is not much unlike a closet shut from light, with only some little opening left, to let in external visible resemblances, or ideas of things without: ...[these] resemble the understanding of a man, in reference to all objects of sight and the ideas of them."

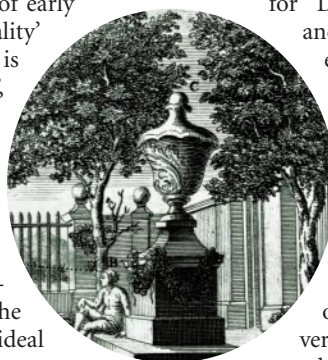
Here we have the birth of early modern epistemology: 'reality' is 'external', knowledge is 'represented' and 'internal', and 'objective truth' has to be a 'correspondence' between the object and its representation. But with this model of knowledge comes the problem of the inner homunculus or 'subject', the self trapped inside the camera, and the need for an ideal observer who sees both what goes on inside and outside at the same time and is thus able to tell whether the object and its representation correspond. Such is the epistemology produced by the engine of the camera obscura.

This progressive history of an epistemology engine displays two movements associated with it. The first is one of escalation — from Alhazen's observation of an optical effect; to da Vinci's camera as analogue for the eye; to Locke's and Descartes' analogue of camera to eye to mind — by which the

camera is made into a full epistemology engine. The second is the inward progression of the location where 'external' reality, itself an artefact of the geometry of the imaging phenomenon, interfaces with the 'inner' representation. For da Vinci, the interface of external/internal occurs "in the pupil"; for Descartes, it is the retina; and, still continuing the camera epistemology, contemporary neuroscience locates it in the brain.

As well as being an amazingly persistent epistemology engine, the ghost of this forgotten technology lurks in the 'science wars'. At least one dimension of this contemporary controversy revolves around a passionate defence of notions such as 'external reality', 'truth' usually defined as 'correspondence', which is modernism's form of 'objectivity', and so on. Those who have begun to question this epistemology (and its antique engine) are dubbed postmodernists and relativists, the latter of whom have yet to agree upon or produce a new epistemology engine. Perhaps it is time, however, to explore a wider array of possibilities. ■

Don Ihde is in the Department of Philosophy, State University of New York, Stony Brook, New York 11794-3750, USA.



The abdication of Pope Mary III

...or, Galileo's revenge...

Robert J. Sawyer

Darth Vader's booming voice, still the network's trademark 600 years after its founding: "This is CNN."

And then the news anchor: "Our top story: Pope Mary III abdicated this morning. Giancarlo DiMarco, our correspondent in Vatican City, has the details. Giancarlo?"

"Thanks, Lisa. The unprecedented has indeed happened: after 312 years of service, Pope Mary III stepped down today. Traditionally, the conclave of Roman Catholic cardinals waits 18 days after the death of a pope before beginning deliberations to choose a successor, but Mary — who has returned to her birth name of Sharon Cheung — is alive and well, and so the members of the conclave have already been sealed inside the Vatican Palace, where they will remain until they've chosen Mary's replacement.

"Although no new pope has been elected for over 300 years, the traditional voting method will be used. We are now watching the Sistine Chapel for the smoke that indicates the ballots have been burned following a round of voting. And — Lisa, Lisa, it's happening right now! There's smoke coming out, and — no, you can hear the disappointment of the crowd. It's black smoke; that means no candidate has yet received the required majority of two-thirds plus one. But we'll keep watching."

"Thanks, Giancarlo. Let's take a look at Pope Mary's press conference, given earlier today."

Tight shot on Mary, looking only a tenth of her 400 years: "Since Vatican IV reaffirmed the principle of papal infallibility," she said, "and since I now believe that I was indeed in error 216 years ago when I issued a bull instructing Catholics to reject the evidence of the two Benmergui experiments, I feel compelled to step down ..."

"We're joined now in the studio by Joginder Singh, professor of physics at the University of Toronto. Dr Singh, can you explain the Benmergui experiments for our viewers?"

"Certainly, Lisa. The first proved that John Cramer's transactional interpretation of quantum mechanics, proposed in the late twentieth century, is in fact correct."

"And that means ...?"

"It means that the many-worlds interpretation is flat-out wrong: new parallel universes are not spawned each time a quantum event could go multiple ways. This is the one and only extant iteration of reality."



"And Dr Benmergui's second experiment?"

"It proved the current cycle of creation was only the *seventh* such ever; just six other Big Bang/Big Crunch oscillations preceded our current universe. The combined effect of these two facts led directly to Pope Mary's crisis of faith, specifically because they proved the existence of — one might as well use the word — God."

"How? I'm sure our viewers are scratching their heads ..."

"Well, you see, the observation, dating back to the twentieth century, that the fundamental parameters of the universe seem fine-tuned to an almost infinite degree specifically to give rise to life, could previously be dismissed as a statistical artefact caused by the existence of many contemporaneous parallel universes or a multitude of previous ones. In all of that, every possible combination would crop up by chance, and so it wouldn't be remarkable that there was a universe like this one — one in which the force of gravity is just strong enough to allow stars and planets to coalesce but not just a little bit stronger, causing the universe to collapse long before life could have developed. Likewise, the value of the strong nuclear force, which holds atoms together, seems finely tuned, as do the thermal properties of water, and on and on."

"So our universe is a very special place?"

"Exactly. And since, as Kathryn Benmergui

proved, this is the *only* current universe, and one of just a handful that have ever existed, then the life-generating properties of the very specific fundamental constants that define reality are virtually impossible to explain except as the results of deliberate design."

"But then why would Pope Mary resign? Surely, if science has proven the existence of a creator ...?"

Singh smiles. "Ah, but that creator is clearly not the God of the Bible or the Torah or the Qur'an. Rather, the creator is a physicist, and we are one of his or her experiments. Science hasn't reconciled itself with religion; it has *superseded* it, and —"

"I'm sorry to interrupt, Dr Singh, but our reporter in Vatican City has some breaking news. Giancarlo, over to you ..."

"Lisa, Lisa — the incredible is happening. At first I thought they were just tourists coming out of the Sistine Chapel, but they're not — I recognize Fontecchio and Leopardi and several of the others. But none of them are wearing robes; they're in street clothes. I haven't taken my eyes off the chapel: there's been no plume of white smoke, meaning they haven't elected a new leader of the church. But the cardinals *are* coming out. They're coming outside, heading into St Peter's Square. The crowd is stunned, Lisa — it can only mean one thing ..."

Nebula Award-winner Robert J. Sawyer's latest novel is *Calculating God (Tor)*; <http://www.sfwriter.com>.

Schrödinger's cat is now fat

Gianni Blatter

Schrödinger's dead-and-alive cat was a thought experiment applying the physics of electrons and atoms to our macroscopic world. New experiments with superconductors narrow the gap between theoretical ideas and reality.

Physics at the atomic scale is determined by the rules of quantum mechanics, which tell us that particles propagate like waves, and so can be described by a quantum-mechanical wavefunction. As an immediate consequence, a particle can be in two or more states at the same time — a so-called superposition of states. This curious behaviour has been hugely successful in describing physical systems at the microscopic level. For example, under the rules of quantum mechanics two atoms sharing an electron form a chemical bond, whereas in classical theory the electron remains confined to one atom and the bond can't form.

Applying quantum theory to our macroscopic world, Schrödinger¹ proposed a *Gedanken* or thought experiment with an unfortunate cat lingering in the twilight zone between life and death. This raised questions about whether quantum mechanics breaks down when the system is complex enough. On page 43 of this issue, Friedman *et al.*² go some way to answering this fundamental question by confirming that quantum superposition works as well in the macroscopic world of superconducting rings as it does in the microscopic world of photons, electrons and atoms.

Discussions about the boundary separating the microscopic quantum and macroscopic classical worlds sharpened up in the early 1980s, when Leggett and others suggested that a macroscopic system could behave quantum mechanically if it was suitably decoupled from its environment³. One of the systems under discussion was a superconducting quantum interference device (SQUID). This is a superconducting ring in which billions of paired-up electrons move in perfect harmony and without resistance. The question was whether a 'persistent' electric current in the ring would decay in a quantum-mechanical way (Fig. 1a). Subsequent experiments showed that currents in SQUIDs did indeed decay by a quantum-mechanical tunnelling process.

If a macroscopic system can decay by quantum tunnelling, the next question to ask is whether it could also oscillate back and forth between two states (Fig. 1b–d), say two states defined by currents flowing clockwise and anticlockwise in a SQUID. And if it can, would the macroscopic state tunnel back and forth sequentially, forgetting about its quantum-mechanical make-up after each

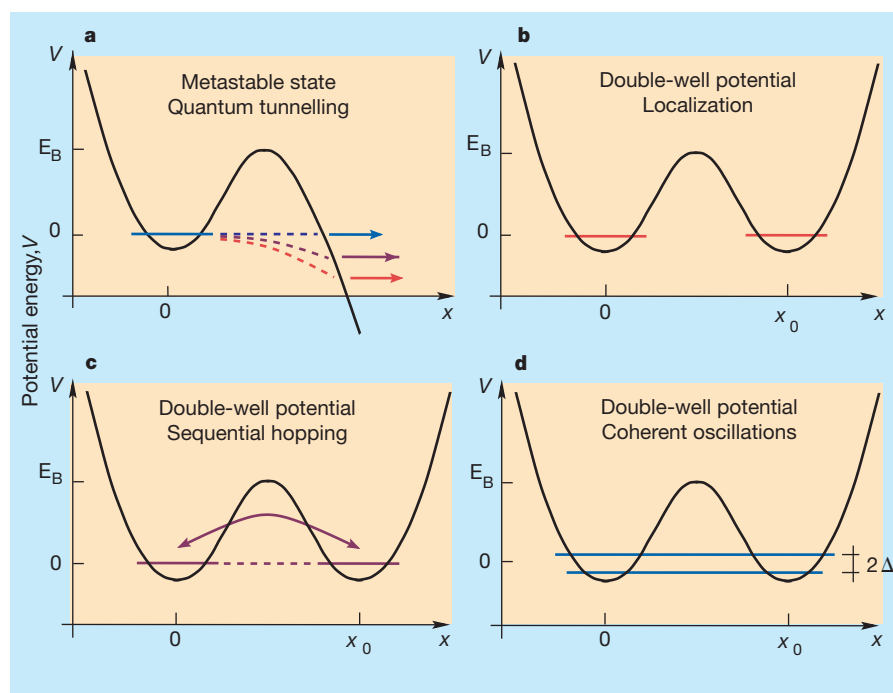


Figure 1 The state of Schrödinger's cat. Physicists use potential wells to describe the preferred states of a system (the positions of electrons in a diatomic atom, the opposing currents in a superconducting ring, or the dead-and-alive state of Schrödinger's cat). A classical system trapped at the bottom of the well remains there because it needs a lot more energy to get out of the well. But a quantum particle trapped in a metastable state can tunnel out of the potential well (a). No energy is lost in the tunnelling process if the particle is isolated from its environment (blue), whereas coupling to the environment means the particle loses energy while tunnelling (weak coupling, magenta; strong coupling, red). b–d, A particle trapped in a double-well potential is localized in one of the wells if its coupling to the environment is large (b), if it hops incoherently back and forth between the wells at intermediate coupling (c), or if it coherently oscillates between the two wells with weak environmental coupling (d). The ground state of the quantum particle in d is lower because extending the wavefunction over both wells allows the particle to lower its energy; the first excited state comes to lie above zero energy. This is the origin of the coherence gap, 2Δ , successfully measured by Friedman *et al.*¹ in their superconducting device. (E_B is the height of the energy barrier.)

hop, or would it oscillate coherently, preserving its quantum state throughout the hops? The second process goes under the name of 'macroscopic quantum coherence' and has been something of a holy grail in this field since the 1980s — it is the coherent superposition between the two current states in the ring that corresponds to the indeterminate state of Schrödinger's dead-and-alive cat.

Quantum theory predicts^{4,5} that if such a system is strongly coupled to the environment, it remains localized in one state and so behaves classically (Fig. 1b). For intermediate coupling, the system hops randomly back and forth between the two states (Fig. 1c). And finally, at low coupling, the system follows

damped, coherent oscillations between the states, with the damping rate vanishing as the coupling to the environment goes to zero. This quantum mixing of the two states leads to a so-called coherence gap (2Δ) separating the energies of the superposition states (Fig. 1d).

One explanation for this coherence gap is as follows. Consider a particle trapped in a potential well with two minima (a double-well potential). As the particle tunnels between the two wells, it lowers its kinetic energy because of the spreading of its wavefunction over both wells. As a result, the new mixed ground state is shifted down by Δ with respect to the energy of the individual wells. This 'symmetrical' state always comes with

an 'antisymmetric' partner state, which is slightly higher in energy. In fact, this excited state comes to lie at an energy Δ above the original well energy, resulting in an excitation gap of 2Δ . This phenomenon is well known in chemistry, where the two mixed or superposition states correspond to the bonding and antibonding states of a diatomic molecule.

Measuring this coherence gap in a macroscopic device, such as a SQUID, is a tricky business because the quantum superposition is extremely sensitive to external noise. Experimentalists have been trying to do this for years with no conclusive results. In their experiment, Friedman *et al.*² use a SQUID with two current states flowing clockwise and anticlockwise. Starting with a specific current state, say clockwise, they then bathe the SQUID in microwaves and measure the probability of finding the current flowing in the opposite — that is, anticlockwise — direction. This probability peaks when the microwave frequency matches up with the two superposition states, allowing them to identify a coherence gap of ~ 0.01 millielectron volts (corresponding to a temperature measurement of 0.1 Kelvin). The size of the gap puts an upper limit on the decoherence rate of the device — that is, the rate at which it decays into a classical state — and allows them to assess the quality of their experimental set-up.

How does this experiment achieve what was previously impossible? Rather than using the ground states of their potential wells, Friedman *et al.* use high-energy microwaves to shift their coherence experiment up to energy states close to the top of the potential barrier, thereby increasing the tunnelling rate between the two wells by many orders of magnitude. The challenge in future experiments will be to track the appearance of this coherence gap when probing lower-energy (semiclassical) states deeper in the well. Another possibility would be to make the coherence gap vanish by introducing artificial decoherence into the system, for example by coupling the SQUID to a metallic reservoir.

So quantum theory does not break down when the system becomes more complex — here, there are billions of electron pairs acting in unison. Are there reasons to suspect that it will break down elsewhere? Most physicists would bet that quantum theory is safe, and that, where it does 'break down', this is because of coupling to the environment inducing decoherence. Besides these philosophical implications, the experiment of Friedman *et al.* has practical relevance, particularly in the newly emerging field of quantum computing. For physical systems to operate as quantum bits and logic gates they have to deal with two conflicting requirements: on one hand, the quantum bits have to be manipulated in

order to carry out the operations required by the computing algorithm; on the other hand, the whole quantum computer must be decoupled from the environment as much as possible in order to preserve its quantum coherence.

Proposals for quantum bits put forward by the quantum optics community are usually based on individual photons or atoms, and have the advantage of being decoupled from the environment. More recently, solid-state systems based on nano-engineered superconducting devices have been proposed as quantum bits, with the advantage that they are easy to manipulate. The first 'superconducting-charge qubit' was built last year by Nakamura *et al.*⁶ They reported real-time observations of coherent charge oscillations on a small superconducting island. But as with coherence experiments on single photons or atoms, this experiment

deals with the coherent propagation of individual electron pairs. What is truly amazing about the new experiment of Friedman *et al.*² is that the tunnelling object is a magnetic flux set up by billions of electron pairs coherently circulating within a superconducting ring. With the discovery of coherence in such a macroscopic system, Schrödinger's cat has put on quite a bit of weight. ■

Gianni Blatter is at the Institut für Theoretische Physik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland.

e-mail: blatterj@itp.phys.ethz.ch

1. Schrödinger, E. *Naturwissenschaften* **23**, 807–812, 823–828, 844–849 (1935).
2. Friedman, J. R., Patel, V., Chen, W., Tolpygo, S. K. & Lukens, J. E. *Nature* **406**, 43–46 (2000).
3. Caldeira, A. O. & Leggett, A. J. *Ann. Phys.* **149**, 374–456 (1983).
4. Schmid, A. *Phys. Rev. Lett.* **51**, 1506–1509 (1983).
5. Chakravarty, S. & Leggett, A. J. *Phys. Rev. Lett.* **52**, 5–8 (1984).
6. Nakamura, Y., Pashkin, Yu. A. & Tsai, J. S. *Nature* **398**, 786–788 (1999).

Signal transduction

A cellular rescue team

Joel L. Pomerantz and David Baltimore

The cells that make up multicellular organisms depend on molecular circuitry designed to receive and interpret signals from the environment, and respond appropriately. Signals from neighbouring cells or tissues can instruct a particular cell to proliferate, differentiate into another cell type, or even commit suicide by programmed cell death. So the accurate functioning of the signalling circuitry is critical for normal development and physiology. Errors can lead to uncontrolled cell growth and diseases such as autoimmunity and cancer.

Hoeflich and colleagues¹, writing on page 86 of this issue, have identified a new player in the cell-survival circuitry. This protein, glycogen synthase kinase-3 β (GSK-3 β), was already known for its involvement in early embryonic development and cell-fate decisions in flies, mammals and frogs. Now GSK-3 β can be ascribed a completely separate function — in a cell's decision to live or die.

The key signalling pathway here is that beginning with tumour-necrosis factor- α (TNF- α ; Fig. 1). This is a proinflammatory cytokine — a signal released during infection that calls the immune system to action. TNF- α activates two signalling pathways within cells. One leads to programmed cell death (apoptosis), whereas the other counters the death signal and leads to survival. When both survival and death pathways are activated by TNF- α , the cell receiving the signal stays alive and multiplies. But if the balance is somehow shifted to the death pathway, it's curtains for the cell. The survival pathway activates a transcription factor, NF- κ B, which works by

turning on a set of anti-apoptotic genes. The intriguing finding made by Hoeflich *et al.*¹ is that GSK-3 β is required for the NF- κ B-mediated survival response.

NF- κ B is normally composed of two subunits, p50 and p65. It is usually held captive in the cytoplasm of a cell because it associates with an inhibitor protein called I κ B (Fig. 1), which stops NF- κ B from entering the nucleus. But when cells are treated with TNF- α , the I κ B protein becomes labelled with phosphate groups, which mark it out for degradation. With its jailer destroyed, NF- κ B is free to move into the nucleus, where it binds to relevant sites in its target genes and activates a new programme of gene expression, ensuring that the cell survives. The I κ B kinase complex mediates the key phosphorylation of I κ B in this chain of events.

The importance of NF- κ B is shown by the fact that, when it is inhibited during cell stimulation with TNF- α , the death pathway is not countered by anti-apoptotic factors, so the cell dies^{2–5}. In the developing mouse, this act is played out in the liver. Mice lacking p65 die during embryogenesis at days 14 to 15 with massive apoptosis of hepatocyte liver cells⁶. But p65-deficient mice that also lack TNF- α ⁷ or one of its receptors (TNFR1) survive. Similar defects are observed in mice deficient for either a catalytic subunit^{8–10} or a non-catalytic subunit¹¹ of the I κ B kinase complex. Embryonic fibroblast cells from these mice have defects in NF- κ B activation and heightened sensitivity to TNF- α -mediated apoptosis.

Hoeflich *et al.*¹ have now generated GSK-3 β knockout mice and, strikingly, these mice



100 YEARS AGO

When the pseudopodium of an *Amoeba* has reached a certain development it suddenly retracts, or rather collapses, for [in this book] Kassowitz regards the phenomenon as a rapid tumbling to pieces of the molecular structure owing to stimulation: certain protoplasm-molecules are shattered, atom-groupings of carbon and hydrogen split the molecular oxygen and are at once burnt to CO_2 and OH_2 , the heat-vibrations evolved during the combustion shattering more molecules, and so on, throughout that part of the mass. This process exhausted, a period of restitution sets in, and new molecules are built up from the fragments of proteids, carbohydrates, fats and mineral substances at disposal, and become interpolated between those which had escaped destruction, and a new pseudopodium is put out by assimilative growth. Among other arguments for the view that this is really a process of growth, Kassowitz points out that the rate of protrusion of such a pseudopodium, rapid as it appears under a high power, is really not much more rapid than the growth of a stem of asparagus, a mushroom or a bamboo.

From *Nature* 5 July 1900.

50 YEARS AGO

It is the purpose of this communication to describe a novel chromatographic method of using ion-exchange resins so that only one of two possible adsorption mechanisms associated with their use can function. Weak organic electrolytes and, in particular, aromatic organic ions can be adsorbed on an ion-exchange resin by a combination of salt linkages and Van der Waals' adsorption forces, if the electrical charge of the resin has the opposite sign to that of the organic ions. Partridge and Davies have recently... shown that a major limitation of ion-exchange resins to-day is their inability to fractionate mixtures containing aromatic molecules. It is suggested that these difficulties can be resolved if aromatic molecules are adsorbed on a dissociated ion-exchange resin the electrical charge of which has the same sign as the aromatic ions. Under these conditions only the undissociated aromatic molecules can be adsorbed, since the corresponding ions cannot approach the surface against its repulsive electrostatic forces. Any change in the pH which dissociates the adsorbed aromatic molecules will result in their desorption.

D. E. Weiss

From *Nature* 8 July 1950.

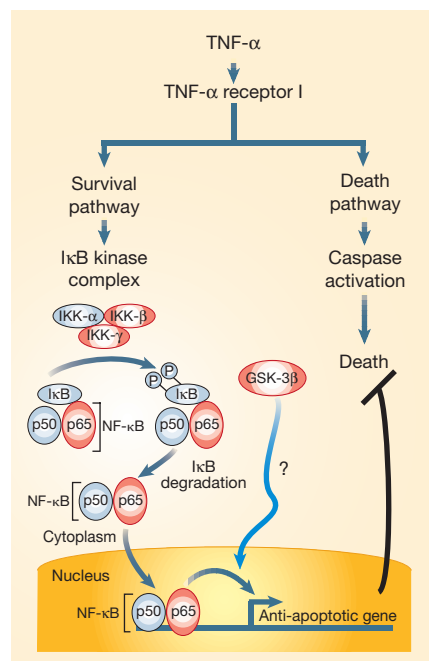


Figure 1 Life-and-death decisions in cells. Cellular stimulation with tumour-necrosis factor- α (TNF- α , top) simultaneously activates survival (left) and death (right) signalling pathways. The survival pathway leads to the activation of NF- κ B, which induces the expression of anti-apoptotic genes in the nucleus. NF- κ B (subunits p50 and p65) is normally held captive in the cytoplasm by the I κ B protein. Cell stimulation with TNF- α leads to activation of the I κ B kinase (IKK) complex, which phosphorylates I κ B. The phosphate tag (circled 'P') singles out I κ B for destruction. NF- κ B is then free to move into the nucleus and activate its target genes. Hoefflich *et al.*¹ have revealed an unexpected requirement for glycogen synthase kinase-3 β (GSK-3 β) in the NF- κ B-mediated activation of genes needed for survival. It is not yet clear how GSK-3 β works in this pathway, but it probably involves a critical step following the movement of NF- κ B to the nucleus. Targeted disruption in mice of any of the molecules coloured red leads to death of the embryo, accompanied by TNF- α -induced apoptosis of hepatocytes.

suffer from defects similar to those of mice lacking components of the NF- κ B pathway. GSK-3 β -deficient mice are morphologically normal at day 12 of embryonic development, but die between days 13.5 and 14.5 with hepatocyte apoptosis, which can be prevented by injection of antibodies that block the function of TNF- α . Embryonic fibroblast cells from these mice are overly sensitive to TNF- α -mediated apoptosis, but not to other apoptotic stimuli. In addition, GSK-3 β -deficient fibroblasts show reduced induction of an NF- κ B-responsive 'reporter' gene in response to TNF- α or interleukin-1, another cytokine involved in immune responses.

Intriguingly, the GSK-3 β -deficient cells still allow the degradation of I κ B and the

release of its prisoner, NF- κ B, to the nucleus. Presumably GSK-3 β is required for a later step in the signalling pathway (Fig. 1) — perhaps once NF- κ B has already bound to its target genes. One possibility is that GSK-3 β is directly or indirectly required for phosphorylation of the NF- κ B subunit p65. This modification is important for the activation of transcription by NF- κ B¹², and occurs when cells are stimulated with TNF- α or interleukin-1 (refs 13, 14). GSK-3 β does not need to enter the nucleus to achieve this task. Instead, it may influence the modification of p65 in the cytoplasm, before or during the translocation of NF- κ B into the nucleus. But there are other possibilities, too, and we are sure to learn much about how NF- κ B is regulated by finding out how GSK-3 β works in this rescue response.

Interestingly, I κ B is itself encoded by an NF- κ B-responsive gene, and is induced as part of a negative feedback loop to inhibit NF- κ B activity. It appears from the results of Hoefflich *et al.* that TNF- α -mediated induction of I κ B expression is intact in GSK-3 β -deficient cells. So GSK-3 β may be required for the expression of only a subset of NF- κ B-responsive genes. But further work is needed to discover what determines this specificity of GSK-3 β action.

The results of Hoefflich *et al.* are also remarkable for the effects on embryogenesis that do not occur. GSK-3 β is involved in the Wnt/Wingless signalling pathway¹⁵, which is required for early embryonic development and cell-fate determination. Here, GSK-3 β works together with a tumour-suppressor protein called adenomatous polyposis coli and with a protein called Axin to phosphorylate β -catenin, singling it out for destruction. If GSK-3 β activity is inhibited, β -catenin accumulates in the cytoplasm. It can then enter the nucleus, bind to certain protein co-factors, and help to induce the expression of Wnt/Wingless target genes. Proper control of β -catenin levels is therefore critical for axis determination and other events in embryogenesis. So another surprise in the results of Hoefflich *et al.* is that their GSK-3 β -deficient mice do not suffer morphological defects before day 12 of embryonic development. Perhaps GSK-3 α or other related proteins functionally substitute for GSK-3 β during early development.

Given that uncontrolled accumulation of β -catenin is associated with colon cancer¹⁶, it would be interesting to know the incidence of tumours in mice lacking one or both copies of the gene encoding GSK-3 β . Mice lacking both copies of the GSK-3 β gene die as embryos, but perhaps mice that lack TNF- α or its receptor TNFR1, as well as GSK-3 β , might survive for long enough to make such a study possible.

The unexpected participation of GSK-3 β in this hepatocyte survival pathway highlights the need to consider the biological

importance of interactions between seemingly independent signalling pathways. Will signals that inhibit GSK-3 β activity (such as Wnt, insulin and epidermal growth factor) also inhibit the NF- κ B-mediated activation of subsets of target genes? The report by Hoeflich *et al.*¹ reveals new opportunities for exploring how networks of signalling circuitry control the life and death decisions of cells. ■

Joel L. Pomerantz and David Baltimore are in the Division of Biology, California Institute of Technology, 1200 E. California Boulevard, MC 20431, Pasadena, California 91125, USA.
e-mails: jpomeran@caltech.edu
baltimo@caltech.edu

1. Hoeflich, K. P. *et al.* *Nature* **406**, 86–90 (2000).
2. Beg, A. A. & Baltimore, D. *Science* **274**, 782–784 (1996).

3. Wang, C. Y., Mayo, M. W. & Baldwin, A. S. Jr *Science* **274**, 784–787 (1996).
4. Van Antwerp, D. J., Martin, S. J., Kafri, T., Green, D. R. & Verma, I. M. *Science* **274**, 787–789 (1996).
5. Liu, Z. G., Hsu, H., Goeddel, D. V. & Karin, M. *Cell* **87**, 565–576 (1996).
6. Beg, A. A., Sha, W. C., Bronson, R. T., Ghosh, S. & Baltimore, D. *Nature* **376**, 167–170 (1995).
7. Doi, T. S. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 2994–2999 (1999).
8. Li, Q., Van Antwerp, D., Mercurio, F., Lee, K. F. & Verma, I. M. *Science* **284**, 321–325 (1999).
9. Tanaka, M. *et al.* *Immunity* **10**, 421–429 (1999).
10. Li, Z. W. *et al.* *J. Exp. Med.* **189**, 1839–1845 (1999).
11. Rudolph, D. *et al.* *Genes Dev.* **14**, 854–862 (2000).
12. Zhong, H., Voll, R. E. & Ghosh, S. *Mol. Cell* **1**, 661–671 (1998).
13. Wang, D. & Baldwin, A. S. Jr *J. Biol. Chem.* **273**, 29411–29416 (1998).
14. Sizemore, N., Leung, S. & Stark, G. R. *Mol. Cell. Biol.* **19**, 4798–4805 (1999).
15. Dale, T. C. *Biochem. J.* **329**, 209–223 (1998).
16. Barker, N., Morin, P. J. & Clevers, H. *Adv. Cancer Res.* **77**, 1–24 (2000).

Theoretical physics

Why trapped atoms are attractive

James Anglin

Before Newton postulated the inverse square law for gravitational force, scientists were already familiar with the inverse square law for light intensity as a function of distance from the source. Today, physics lecturers often introduce students to the gravitational inverse square by comparing it to that of light. A paper from O'Dell *et al.* in *Physical Review Letters*¹ suggests that we can now reverse this old analogy, and induce an inverse square attractive force between particles using light itself. This will allow experiments that simulate the gravitational interactions of large numbers of particles. And it illustrates a current trend in the physics of atoms confined in traps — experimental control has become sufficiently strong that one can make clouds of cold, trapped atoms that emulate a surprisingly wide range

of physical systems, so we can investigate the physics behind these phenomena under tightly controlled conditions (Fig. 1).

Experiments with ultracold atoms have accelerated and diversified greatly since the creation of the first Bose–Einstein condensate in 1995. Condensates are samples of ultra-cold atoms that share the same quantum state and therefore have remarkable properties. Although the 1995 experiments were impressive breakthroughs in technology, from the viewpoint of fundamental theory, Bose–Einstein condensation itself was 70-year-old news. Today, however, condensates can be made to do things Bose and Einstein never imagined.

This versatility is what makes condensates so exciting, and which allows us to do things that were previously impossible, such

as tuning the interactions between particles. Atomic properties can be changed in many ways, by applying laser beams or magnetic fields, but previous techniques were limited to strengthening, weakening or possibly reversing the natural short-range forces between atoms.

O'Dell *et al.*¹ now propose bathing a cloud of trapped atoms in laser light to induce a long-range attractive force between each and every other atom. A static electric field will induce a force between atoms that decays with the inverse fourth power of their separation. But in the oscillating field of a light beam, there is an additional force that falls away as the inverse square. The sum of these forces may be either attractive or repulsive for a given pair of atoms, depending on the angles between the atoms and the electric field. But by shining a whole battery of lasers on the sample from many directions, one can effectively average over all angles, to give a net attraction that is purely inverse square. (The need for so many lasers is a practical obstacle to carrying out such an experiment, but it will hardly be a permanent barrier.)

To prevent the laser fields from simply destroying the cloud of trapped atoms, the laser wavelengths must be much longer than the size of the sample, and their frequencies must be far from any resonant frequencies of the atoms. Under these conditions one can safely use intense laser fields to create a strong artificial 'gravity' between the particles that is many orders of magnitude greater than their genuine gravitational attraction.

The authors show that a Bose–Einstein condensate subjected to this artificial gravity-like interaction will exhibit interesting phase transitions at low temperature, and may even remain bound together under its own 'gravity', with the external trapping field turned off. (In this case, 'normal' gravity limits the duration of the experiment, as the atomic cloud falls out of the laser field. But the US space agency NASA is already

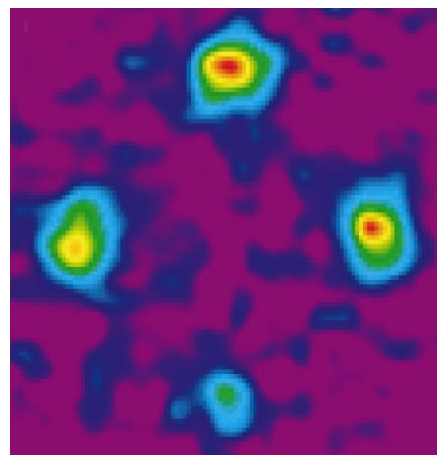


Figure 1 Trapped atoms as a Rosetta stone. Theorists can now dream of using ultracold matter (left, from ref. 11) to model physical systems ranging from black holes to stellar clusters (right). O'Dell *et al.* suggest that, with the help of a few lasers, a cloud of ultracold atoms can even be made to simulate gravity.

considering proposals to run trapped-atom experiments in orbit, to avoid such problems.) This self-bound state will allow researchers to simulate aspects of stellar physics that are difficult to examine in the laboratory. And one can expect further possibilities in this direction, such as simulations of galaxies and star clusters.

In the past year, there have been several other proposals to use light fields to persuade cold, trapped atoms to model the physics of quite different systems. Bose–Einstein condensate analogues have been suggested for cosmic strings (structural defects) formed in the early Universe², superconductor–insulator phase transitions³ and Josephson oscillations⁴ in solid-state systems, and general relativistic black holes⁵. The reason such schemes are now thinkable is that the necessary experimental techniques have become available. Experimenters can choose the way their atoms interact⁶, project them in coherent beams⁷, load them into optical lattices⁸, and efficiently ‘engineer’ them into exotic collective states^{9,10}. Decades of technical work are now yielding a profound control over matter and light. Ingenious simulations, such as that proposed by O’Dell *et al.*, will be useful benchmarks as this technology advances further, as well as being one of its most promising applications.

There is no way that millions of trapped atoms — interacting quantum mechanically and nonlinearly — will merely duplicate a theoretician’s model, so the proposed models are genuine experiments. But they are extremely controlled experiments in which

we can examine the principles involved in complex phenomena in a greatly simplified medium. Such highly controlled analogues can serve as Rosetta stones, in which (as it were) the message of an undeciphered script appears duplicated in a familiar language, so that we can begin translating. One hopes that this basic strategy will be extended to more and more scientific questions as techniques improve.

There will, of course, be vital aspects of many natural phenomena that will simply be lost in this kind of simplification. One will not find the keys dropped in a dark alley under the lamppost just because the light is stronger there. But it is worth trying to shine the light a bit further, as it seems we now can, because we can see so much more in strong light. We can even see gravity. ■

James Anglin is at the Institute for Theoretical Atomic and Molecular Physics, 60 Garden Street, MS 14, Cambridge, Massachusetts 02138, USA.
e-mail: janglin@cfa.harvard.edu

1. O’Dell, D., Giovanazzi, S., Kurizki, G. & Akulin, V. M. *Phys. Rev. Lett.* **84**, 5687–5690 (2000).
2. Anglin, J. R. & Zurek, W. H. *Phys. Rev. Lett.* **83**, 1707–1710 (1999).
3. Giovanazzi, S., Smerzi, A. & Fantoni, S. *Phys. Rev. Lett.* **84**, 4521–4524 (1999).
4. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. *Phys. Rev. Lett.* **81**, 3108–3111 (1999).
5. Leonhardt, U. & Piwnicki, P. *Phys. Rev. Lett.* **84**, 822–825 (2000).
6. Stenger, J. *et al.* *Nature* **396**, 345–348 (1998).
7. Hagley, E. W. *et al.* *Science* **283**, 1706–1709 (1999).
8. Anderson, B. P. & Kasevich, M. A. *Science* **282**, 1686–1689 (1998).
9. Matthews, M. R. *et al.* *Phys. Rev. Lett.* **83**, 2498–2501 (1999).
10. Madison, K. W., Chevy, F., Wohlleben, W. & Dalibard, J. *Phys. Rev. Lett.* **84**, 806–809 (2000).
11. Deng, L. *et al.* *Nature* **398**, 218–220 (1999).

ble and UV-A radiation and varying levels of UV-B. Ries *et al.* analysed the recombination frequency in somatic (non-reproductive) cells in response to a natural spectral balance. They also looked at the recombination frequency when UV-B was raised to higher but ecologically relevant levels, and in the presence of a high level of UV-B radiation exceeding that caused by ozone depletion.

The plants showed somatic recombination in response to ecologically relevant increases in UV-B radiation; large elevations in UV-B radiation caused further genomic instability. These results should be regarded as further data points enabling us to interpolate, rather than extrapolate, the mutation rates that may occur in response to increases in the Earth’s UV-B exposure. Irradiation of a mutant *Arabidopsis* line that was defective in the repair of cyclobutane pyrimidine dimers (the major type of DNA damage induced by UV-B exposure) caused further increases in DNA recombination. This observation indicates that recombination might be involved in elimination of pyrimidine dimers.

Reis *et al.*² looked only for a certain type of mutation — the recombination between closely spaced tandem repeats — which could reflect recombination within chromosomes or between replicated copies of chromosomes. It will be interesting to determine whether other types of mutation, such as those occurring at a single DNA base pair, are also increased. The events observed here may represent the tip of a mutational iceberg.

How damaging is this level of mutation to plants in the long term? Most of the recombination events observed were somatic², occurring in tissues that are not going to transmit their genes to the next generation. Such mutations in animals can be lethal, as a cell carrying a single DNA mutation can eventually develop into a tumour. Plants do develop tumours, but plant cells, trapped within their cell walls, are immobile so the tumours cannot spread. It seems, then, that the toxicity induced by rare mutations, including cell-lethal mutations, is insignificant compared with the direct effects of ultraviolet radiation on plant physiology and morphology. But the authors also observed a small number of completely blue progeny of ultraviolet-exposed plants, indicating that recombination had also occurred in the reproductive (germ) cells. So raised UV-B levels may cause permanent changes in plant populations.

These experiments on germinal mutation rates need to be repeated before their significance can be interpreted. But, in this respect, it is interesting that the effects of UV-B on genomic instability increased with each generation; that is, the progeny plants exposed to high UV-B levels had a higher somatic mutation rate than their predecessors. This suggests that the plants are under-

Plant biology

An unbearable beating by light?

Anne B. Britt

The tanning craze of the 1960s and 1970s led to premature ageing and skin cancer in the subsequent decades, and revealed that ultraviolet (UV) radiation causes DNA damage. Awareness of the carcinogenic effects of the UV-B (wavelength 280–320 nanometres) and UV-A (320–400 nanometres) components of sunlight has led to increased concern over the depletion of the Earth’s ozone layer. The relationship between exposure to sunlight, genetic damage and the induction of skin cancer in humans is well documented¹. Increased exposure to UV-B is also known to cause a variety of physiological and morphological responses in plants. But the effect of solar ultraviolet radiation on plant genomic stability has not been well established. On page 98 of this issue, however, Ries *et al.*² report that UV-B radiation can induce DNA rearrangements in the model experimental

plant *Arabidopsis* and in tobacco plants.

The authors looked at the frequency of a particular type of genomic rearrangement that occurred in response to UV-B exposure in transgenic plant lines. These plants carried a ‘reporter’ gene that causes plant cells to stain blue wherever DNA rearrangements between short, closely spaced repeated DNA sequences (recombination) occurred². Because each plant has millions of cells, this system makes it possible to see even very rare recombination events that occur only once per million cells.

The ratio of UV-B to visible light in sunlight changes with altitude and latitude, and the physiological effects of UV-B on plants are influenced by the overall spectrum of radiation, including visible light³. So, to regulate and monitor light exposure, the authors used sophisticated solar simulators. Plants were grown at constant levels of visi-

going heritable and cumulative changes in the expression of genes involved in DNA metabolism. Acute doses of UV-B (as well as ionizing radiation) have been shown to activate transposable elements⁴, and the damage done to DNA by such elements might also be involved in the induction of repair genes, creating a sort of damage–repair chain reaction. It is possible that some sort of signal regarding genomic stress is passed from one generation to the next. This may be an example of what Barbara McClintock described as “the response of the genome to shock”⁵ after observing the activation of transposable elements in maize after the induction of a chromosome break. Whether the environmentally induced genomic instability seen here represents the same or a related phenomenon remains to be seen.

This study presents intriguing data suggesting that depletion of the ozone layer may have a measurable effect on the mutation rate of at least some plants. But further work to determine the mutation rate and the spectrum of mutations caused by UV-B exposure are required before we can fully assess the significance of these observations. ■

Anne B. Britt is at the Section of Plant Biology, University of California, Davis, California 95616, USA.

e-mail: abbritt@ucdavis.edu

1. Brash, D. E. *Trends Genet.* **13**, 410–414 (1997).
2. Ries, G. *et al. Nature* **406**, 98–101 (2000).
3. Rozema, J. *et al. Trends Ecol. Evol.* **12**, 22–28 (1997).
4. Walbot, V. *Nature* **397**, 398–399 (1999).
5. McClintock, B. *Science* **226**, 792–801 (1984).

Aquatic ecology

Phosphorus, the staff of life

David M. Karl

All organisms require nutrients in trace or comparatively much larger (macro) amounts. Phosphorus is one such macronutrient, accounting for about 2–4% of the dry weight of most cells. Life is truly built around phosphorus¹. Yet despite its well-accepted ecological role, quantitative studies of the nutrient's dynamics are limited by the analytical methods available. In many aquatic environments, phosphate — PO_4^{3-} , the preferred substrate for microbial growth — is ‘undetectable’. That is, it is technically below the limit of approximately 30 nanomolar (10^{-9} molar) of standard methods.

On page 54 of this issue, Hudson *et al.*² reveal just how small the ambient pools of phosphate can be. They have assessed phosphate and total phosphorus (both dissolved and particulate forms) in a diverse set of 56 North American lakes of different nutrient status. They find that phosphate levels are two to three orders of magnitude lower than previously thought — the range was from 27 to 885 picomolar (10^{-12}), with a mean of 174 pM. Using an indirect steady-state radio-bioassay, the authors estimated ambient phosphate by direct measurements of the rates of uptake by microorganisms and recycling back

into the ecosystem. Application of the commonly used but imperfect ‘soluble reactive P’ method and the improved ‘Rigler bioassay’ to these same habitats yielded phosphate estimates that were 100–1,000 times higher.

Incredibly, even at ambient phosphate concentrations of 50 pM or less, microbiological uptake in the lake ecosystems was rapid and efficient. This indicates that phosphorus was the limiting nutrient, its availability being the main brake on primary productivity. A counterintuitive result was the significant positive correlation between phosphate and total phosphorus concentrations, implying that low standing stocks of microbes are most efficient at removing phosphate from the water. However, the ambient pool is ultimately controlled by a complex balance of factors — phosphate delivery to and export from the ecosystem; microbial and non-biological uptake and release; and local mineralization of organic phosphorus. One unstated implication of the low phosphate levels and rapid (< 10 min) turnover times in these natural aquatic habitats is that phosphorus flux, not quantity, is probably the more ecologically relevant parameter.

Among the other implications is the

Molecular nanostructures

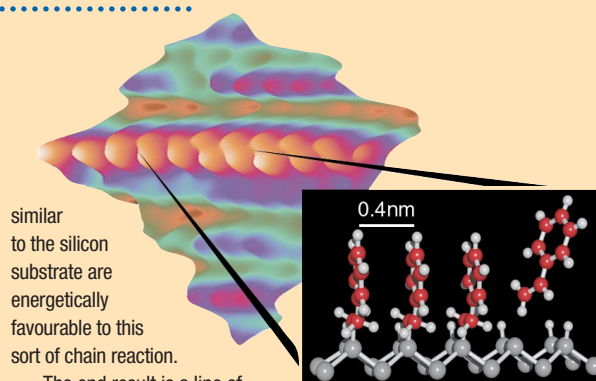
Growth in the fast lane

Controlling the formation of nanostructures at the molecular scale is a great technological challenge in the design of electronic circuits and devices. Molecular manipulation is a tricky business and, although there is scope for greater miniaturization of electronic components, future developments will require an even closer integration between chemistry, biology, physics and engineering. The most promising approaches so far — using scanning tunnelling microscopes (STM) or self-assembly processes — are either too slow or do not allow accurate monitoring of the growth of these structures.

Elsewhere in this issue (*Nature* **406**, 48–51; 2000), Robert Wolkow and colleagues describe how they have created one-dimensional organic structures on a silicon surface, by combining STM and self-directed growth techniques. The picture on the left shows an STM image of a straight line of styrene

molecules (C_8H_8 ; red) attached to a silicon surface. This is an ideal substrate for the assembly of straight ‘wires’ because it is anisotropic; that is, the growth process proceeds in one direction, which depends on the crystallographic orientation of the surface.

How are these molecular wires produced? The first step is to expose a clean surface of silicon to atomic hydrogen in an ultra-high vacuum to obtain a surface terminated by hydrogen atoms. A hydrogen atom is then removed from the surface with the tip of an STM to create a single silicon dangling bond, which will then spontaneously react with any available styrene to form a silicon–carbon bond. To compensate for this newly reactive carbon species, it is likely that a hydrogen atom is removed from a neighbouring surface site, leading to a chain reaction. Theoretical calculations have shown that anisotropic surfaces



similar to the silicon substrate are energetically favourable to this sort of chain reaction.

The end result is a line of styrene molecules bonded to the surface. The longest chains created were 13 nm long. The phenyl rings of the styrene molecules were separated by 0.4 nm (right image), which corresponds to the intermolecular distance between two hydrogen atoms on this surface. Although it is not possible to tell whether the rings are actually parallel, some degree of intermolecular coupling between adjacent rings would be necessary for these chains to function as molecular wires.

As the authors point out, the process could be applied to other organic molecules, such as alkenes and alkynes, which may be used to design conducting wires. Many factors need to be explored to help unravel the exact growth mechanism, but this approach is already sufficiently under control for one to imagine the rapid formation and connection of many nanostructures in parallel.

Vincent Dusastre

apparent ability of planktonic microbes to take up phosphate at levels of less than one part per trillion, which appears to be comparable to the assimilation efficiency of many trace elements. At the lowest concentrations reported (under 50 pM), uptake could be limited by diffusion, especially at the inferred growth rates in these habitats. This should select for very small organisms with large surface-to-volume ratios. Ironically, on a mass-for-mass basis, small bacterial cells have higher requirements for phosphorus than do larger organisms in the phytoplankton. So the bacteria may be net sinks for phosphorus, rather than efficient remineralizers of the element as is often assumed.

If this is true, then the supply of phosphorus to the weaker photosynthetic competitors must come from other sources, such as dissolved organic phosphorus or direct ingestion of other organisms, or from vertical migration or other adaptations that enable them to tap into different sources of the nutrient. It is equally likely that the competitive microorganisms in these habitats do not exist in our pure culture collections. So we can expect there to be hitherto unknown transport systems or other physiological adaptations for phosphate uptake.

In recent years there has been a renewed debate over which nutrients are the proximate and ultimate rate-limiting factors on primary production in aquatic environments^{3,4}. In both freshwater and marine habitats, the phosphorus and nitrogen cycles are inextricably linked by the stoichiometry of organic-matter production and remineralization. Under certain conditions, however, microorganisms that can fix nitrogen gas decouple this nutrient interdependency and drive the ecosystem to phosphorus-limitation.

For instance, I have hypothesized that within the past two decades an oceanic area, the North Pacific subtropical gyre, has shifted from a nitrogen-controlled to a phosphorus-controlled habitat as a result of climate-driven selection for the nitrogen-fixing cyanobacterium *Trichodesmium*^{5,6}. From Hudson and colleagues' regression plot of soluble phosphate against total phosphorus, one would predict a further reduction in North Pacific phosphate concentrations from the present 20–50 nM to around 100 pM, or less.

Elsewhere, measurements using a new high-sensitivity, high-specificity assay⁷ in the Sargasso Sea in the North Atlantic, another region of increased nitrogen fixation, indicate sub-nanomolar phosphate concentrations (ref. 8, and J. Wu and E. A. Boyle, unpublished results). In these open-ocean habitats, phosphate concentrations may ultimately be controlled by the atmospheric deposition of dust, which supplies the iron needed to sustain maximal activity of the nitrogenase enzyme responsible for nitrogen fixation⁹.

Finally, 'luxury' uptake and internal storage of phosphorus is a well-documented

metabolic trait in the microbial world. If freshwater microorganisms can efficiently remove phosphate at ambient levels of at least 27 pM, and still sustain optimal growth, then any concentration above that level might be considered excess. From this it would seem that most aquatic habitats, including both freshwater and marine ecosystems, should have only trace levels of phosphate and should be populated with small microbial cells that are enriched in phosphorus relative to carbon or nitrogen.

However, this is not generally observed; indeed, phosphorus-controlled populations often have lower P:C and P:N ratios than expected. An alternative mechanism to achieve the low phosphate levels in the lakes examined by Hudson *et al.*² is adsorption onto clay minerals, iron-rich particles or colloidal organic materials. Although they are much less interesting from a physiological perspective, non-biological as well as biological mechanisms of phosphate removal are ecologically relevant.

All in all, we are left with a sobering thought — that the results of Hudson *et al.* mean that a large literature on phosphate

concentrations in aquatic ecosystems, and also models of nutrient dynamics, may be in jeopardy. The authors do not present data for other macronutrients or trace nutrients. But it seems likely that, at least in the lakes that they studied, phosphorus is the limiting nutrient and so the staff of life — the most essential of nutrients. If the results are confirmed, this study will stand as a pathfinding contribution in aquatic ecology. ■

David M. Karl is in the School of Ocean and Earth Science and Technology, Department of Oceanography, University of Hawaii, Honolulu, Hawaii 96822, USA.

e-mail: dkarl@soest.hawaii.edu

1. de Duve, C. *Blueprint for a Cell: The Nature and Origin of Life* (Patterson, Burlington, NC, 1991).
2. Hudson, J. J., Taylor, W. D. & Schindler, D. W. *Nature* **406**, 54–56 (2000).
3. Falkowski, P. G. *Nature* **387**, 272–275 (1997).
4. Tyrrell, T. *Nature* **400**, 525–531 (1999).
5. Karl, D., Letelier, R., Dore, J., Christian, J. & Hebel, D. *Nature* **388**, 533–538 (1997).
6. Karl, D. M. *Ecosystems* **2**, 181–214 (1999).
7. Thomson-Buldis, A. & Karl, D. *Limnol. Oceanogr.* **43**, 1565–1577 (1998).
8. Cavender-Bares, K. K. *Size Distributions, Population Dynamics, and Single-cell Properties of Marine Plankton in Diverse Nutrient Environments*. Thesis, Massachusetts Inst. Technol. (1999).
9. Michaels, A. F. *et al. Biogeochemistry* **35**, 181–226 (1996).

Population biology

Parasites take control

Richard D. Gregory and Peter J. Hudson

Parasites may make up half of the animal species on Earth¹, yet their impact on the populations of their unwilling hosts is largely unknown². Theory has defined the conditions under which parasites might stabilize and regulate host numbers³. But although it seems that they can cause cyclic fluctuations in host numbers⁴, there is no sound evidence of stable regulation by parasites of natural populations because the experiments required are very difficult.

An alternative method is to look at the relationship between an invading parasite and its host, and show that the pathogen not only reduces the size of the population, but also keeps it low. As they report in *Proceedings of the National Academy of Sciences*⁵, Hochachka and Dhondt have used this approach to demonstrate how parasites can indeed control host numbers.

The authors followed the invasion by a bacterial parasite of a songbird population in eastern North America. The bird — the house finch *Carpodacus mexicanus* — is a small (20 g) native of western North America that was introduced onto Long Island, New York, in 1940. At first the population failed to thrive. But in the 1960s the finch's numbers and range in eastern North America started increasing dramatically, and in 1995 the eastern and western populations met; at the time



Figure 1 A male house finch of the species studied by Hochachka and Dhondt⁵, with the debilitating symptoms of mycoplasmal conjunctivitis. As the disease develops, finches lose weight, become increasingly weak and either die of starvation or are taken by predators.

house finches numbered several millions and occupied most of eastern North America.

The parasite concerned, *Mycoplasma gallisepticum*, causes conjunctivitis in the eyes of infected birds (Fig. 1). It was known to infect poultry, but had not been previously reported in wild birds. The first signs of disease came in 1994, when birdwatchers reported cases of conjunctivitis in the eastern population of finches at garden feeders in Washington DC. The bacterium turned out to be a distinct strain from that found in poultry⁶ and experimental infections have confirmed its

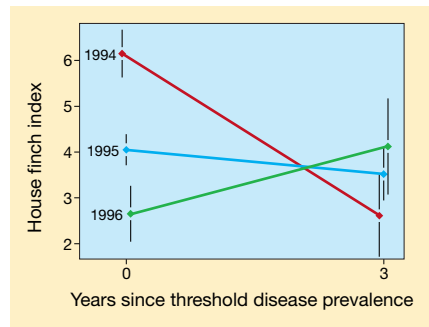


Figure 2 House finches and *Mycoplasma gallisepticum* infection. The plot shows changes in the birds' abundance in three areas where the threshold disease prevalence (20%) was reached in 1994, and the first halves of 1995 and 1996. There were significant differences in finch abundance before the epidemic, but no significant differences three years later. The study⁵ provides convincing evidence that a parasite can regulate host populations in a density-dependent manner.

pathogenicity in captive house finches⁷. Since 1994, an army of volunteers has tracked the spread of the disease month by month across the finch's eastern range^{5,8,9}. To date, the disease has not been reported in the western population.

Armed with information on the proportion of hosts infected and the abundance of the hosts before and after infection, Hochachka and Dhondt showed that there had been a rapid decline in finch numbers at a series of sites with the onset of disease. This effect was most pronounced over the first three years of infection, after which bird numbers stabilized at a lower level. Crucially, bird numbers declined more quickly in those areas infected first, which were also those where the birds were more numerous (Fig. 2) — this is the first demonstration of density-dependent regulation of an animal host by a parasite under natural conditions.

There are many, often spectacular, examples of invading pathogens reducing host abundance (rinderpest in bovine species and phocine distemper virus in seals, for instance¹⁰). Providing evidence of population regulation is far more difficult, however. The strength of the new study⁵ is that the authors had numerous replicate sites where they recorded host abundance before the wave of infection, the time infection arrived and the subsequent host population size.

But there are alternative explanations for the patterns described. For example, the reduced effects of the disease in areas with low bird numbers could be due to declining pathogen virulence. Equally, the course of the disease could have been driven by differences in host susceptibility across the host range. Nevertheless, parasite-induced regulation by a virulent pathogen remains the parsimonious hypothesis. It finds support both from molecular studies that have

shown no change in the *M. gallisepticum* strain through time, and from the observation that chronic disease persists for several years after the initial outbreak¹¹.

Hochachka and Dhondt⁵ also assumed, not unreasonably, that parasite transmission depends on bird density. If infection occurs at feeders, however, then the frequency of feeder use might be a better predictor of transmission than density alone — a possibility not tested by the authors. In reality, the spread of most directly transmitted diseases will be fuelled by a subtle combination of both density- and frequency-dependent processes.

Overall, the story is unusual in several respects. Rarely has a disease been so readily and reliably recognized. Both parasite and host find themselves in the eastern states thanks to respective accidental and deliberate release by human agencies (it is possible that 'founder effects', stemming from a limited gene pool in the initial eastern population, have predisposed these birds to disease).

Moreover, the landscape has been heavily modified by human activity, but house finches are well adapted to such environments, from rural dwellings to suburbia, and take readily to garden feeders. This characteristic, combined with a gregarious nature, has been part of their undoing. Unwittingly, people may have been 'killing with kindness' as feeders increase transmission by bringing birds into close contact, and perhaps by keeping the infectious birds alive longer. More positively, the paper illustrates the potential of 'citizen science'⁸ in harnessing the enthusiasm of volunteers.

The authors now plan to examine the mechanisms of host-parasite interaction inferred from their study. This is a matter of urgency, because the disease is not only still spreading in finches, but has passed into other songbirds such as *Passer domesticus*, *Carpodacus purpureus* and *Carduelis tristis*⁵ — the house sparrow, purple finch and American goldfinch, respectively. ■

Richard D. Gregory is in the Department of Conservation Science, The Royal Society for the Protection of Birds, The Lodge, Sandy, Bedfordshire SG19 2DL, UK.

e-mail: richard.gregory@rspb.org.uk

Peter J. Hudson is in the Department of Biological Sciences, University of Stirling, Stirling FK9 4LA, UK. e-mail: p.j.hudson@stir.ac.uk

1. Price, P. W. *The Evolutionary Biology of Parasites* (Princeton Univ. Press, NJ, 1980).
2. Newton, I. *Population Limitation in Birds* (Academic, London, 1998).
3. Anderson, R. M. & May, R. M. *J. Anim. Ecol.* **47**, 219–247 (1978).
4. Hudson, P. J. *et al. Science* **282**, 2256–2258 (1998).
5. Hochachka, W. M. & Dhondt, A. A. *Proc. Natl Acad. Sci. USA* **97**, 5303–5306 (2000).
6. Ley, D. H. *et al. Emerg. Infect. Dis.* **3**, 375–380 (1997).
7. Luttrell, M. P. *et al. J. Wildl. Dis.* **34**, 289–296 (1998).
8. Dhondt, A. A., Tessaglia, D. L. & Slothower, R. L. *J. Wildl. Dis.* **34**, 265–280 (1998).
9. <http://birds.cornell.edu/hofi/>
10. Dobson, A. P. & Hudson, P. J. in *Ecology of Infectious Diseases in Natural Populations* (eds Grenfell, B. T. & Dobson, A. P.) 52–89 (Cambridge Univ. Press, 1985).
11. Hartup, B. K., Bickel, J. M., Kollias, G. M. & Ley, D. H. <http://www.vet.uga.edu/ivcvm/hartup/hartup.htm> (2000).

Daedalus

Green smelters

Polluted land can be slowly reclaimed by bioremediation — using microorganisms or plants to detoxify the pollution. Certain 'hyperaccumulator' plants can accumulate amazing concentrations of metal in their tissues, often as insoluble or biologically unavailable compounds such as histidine complexes. Plants have no way of excreting toxic metals, and have evolved this defence against them. Metal hyperaccumulators often occur naturally on the metal-rich terrain of mining districts. Daedalus is now taking the biochemistry a step further.

He notes that a bacterial enzyme, mercury reductase, can reduce mercury ions in solution to mercury metal. The gene for this enzyme has recently been transferred to a water-weed. The weed can then reduce deadly mercury ions in the water to the insoluble metal, which simply sits inertly in its tissues. The same enzyme can also reduce silver and cadmium ions. DREADCO biochemists are therefore looking for plant and bacterial enzymes that can work the trick on other metals. They will then insert them into plants which are hyperaccumulators for that metal. The resulting varieties will extract metallic residues from contaminated land, concentrate them, and smelt them internally to pure metal.

Daedalus is not sure what form the metal will take inside the plant. He would like it to occur as one or more big chunks, ideally in the nuts or fruit of the plant, so that it could be picked at harvest time. But it will probably be distributed widely in the cells as tiny single crystals. The plants will have to be reaped in their entirety, and comminuted in water. The suspension could then be centrifuged or sedimented to extract the metal.

This elegant and highly 'green' metal-smelting process will be widely applauded. Unlike conventional smelting, it uses no fuel and puts no CO₂ into the atmosphere. It should be ideal for reclaiming regions despoiled by conventional mining, as well as tracts of low-grade ore too dilute to dig out in bulk. Easily reduced, high-value metals such as silver, copper and nickel are the most obvious targets. Iron looks feasible too — it certainly has a rich biochemistry. Titanium is a tantalizing hope. But the big target is undoubtedly aluminium. It occurs plentifully in almost all soils, and its conventional smelting needs vast amounts of costly electricity. Yet its biochemistry seems forbidding. Only around Chernobyl may plants have mutated enough to have invented the necessary enzymes.

David Jones

Detection of preinvasive cancer cells

Early-warning changes in precancerous epithelial cells can now be spotted *in situ*.

More than 85% of all cancers originate in the epithelium that lines the internal surfaces of organs throughout the body. Although these are readily treatable provided they are diagnosed in one of the preinvasive stages¹, early lesions are often almost impossible to detect. Here we present a new optical-probe technique based on light-scattering spectroscopy that is able to detect precancerous and early cancerous changes in cell-rich epithelia.

Before they become invasive, at stages known as dysplasia and carcinoma *in situ*, early cancer cells alter the epithelial-cell architecture. In particular, the nuclei become enlarged, crowded and hyperchromatic (that is, they stain abnormally darkly with a contrast dye as a result of changes in their chromatin content). These warning signs have so far only been detectable by histological examination of biopsy specimens¹, but light-scattering spectroscopy now offers a biopsy-free means to measure the size distribution and chromatin content of epithelial-cell nuclei as an indicator of preinvasive neoplasia.

The diameter of non-dysplastic cell nuclei is typically 5–10 μm , whereas dysplastic nuclei can be as large as 20 μm across. Epithelial-cell nuclei can be modelled as transparent spheroids that are large in comparison to the wavelength of visible light (0.4–0.8 μm), and whose refractive index is higher than that of the surrounding cytoplasm because of their chromatin content. The spectrum of light backscattered by these particles contains a component that varies characteristically with wavelength², with this variation depending on particle size and refractive index.

For a collection of nuclei of different sizes, the light-scattering signal is a superposition of these variations, enabling the nuclear-size distribution and refractive index to be determined from the spectrum of light backscattered from the nuclei^{3,4}. Once the nuclear-size distribution and refractive index are known, quantitative measures of nuclear enlargement, crowding and hyperchromasia can be obtained.

However, only a small amount of the light incident on the tissue is backscattered by the epithelial-cell nuclei: the rest becomes randomized in direction by multiple scattering, producing a large background of diffusely scattered light that must be subtracted. We have used two tactics to accomplish this — mathematical modelling of the diffusive background³ and polarizing the incident light, which is depolarized by multiple scattering⁴ and enables the single backscattering

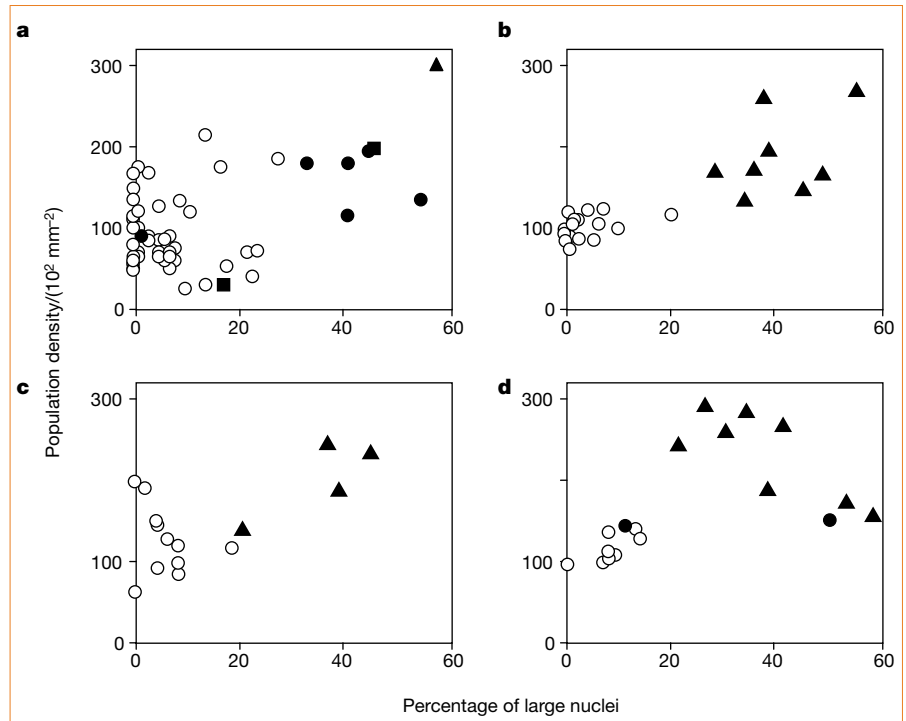


Figure 1 Classification of dysplasia and carcinoma *in situ* by light-scattering spectroscopy compared with diagnosis by histology for four types of epithelial tissue. The population density of nuclei, which determines nuclear crowding, is plotted against the percentage of enlarged nuclei. **a**, Barrett's oesophagus epithelium: circles, non-dysplastic Barrett's mucosa; squares, indefinite diagnosis for dysplasia; filled circles, low-grade dysplasia; triangle, high-grade dysplasia. **b**, Colon epithelium: circles, normal colonic mucosa; triangles, adenomatous polyp. **c**, Urinary bladder epithelium: circles, benign bladder mucosa; triangles, transitional cell carcinoma *in situ*. **d**, Oral cavity epithelium: circles, normal cells; filled circles, low-grade dysplasia; triangles, squamous-cell carcinoma *in situ*. Dysplasia in Barrett's oesophagus was studied during oesophagogastroduodenoscopy in 16 adults with a history of Barrett's oesophagus; colonic dysplasia was studied during colonoscopy in eight adults being screened for adenomatous colon polyps; carcinoma of the urinary bladder was investigated during cystoscopy in seven adults being screened for transitional-cell carcinoma or with a history of haematuria (blood in the urine). Testing for dysplasia and carcinoma in the oral cavity was done during routine oral examination in 5 adults suspected to have squamous-cell carcinoma. Informed consent from the patients was obtained in each case.

to be observed by subtracting the depolarized component of the backscattered light.

We have tested the potential of this technique *in vivo* to diagnose dysplasia and carcinoma *in situ* in four different human organs with three different types of epithelium: columnar epithelia of the colon and Barrett's oesophagus, transitional epithelium of the urinary bladder, and stratified squamous epithelium of the oral cavity. Our clinical investigations were all made during routine endoscopic screening or surveillance procedures, during which an optical-fibre probe delivered white light from a xenon arc lamp to the tissue surface and collected the returning light³.

The tip of the optical probe was brought gently into contact with the tissue under investigation. Immediately after measuring the backscattered light, a tissue biopsy sample was taken from the same site for histological examination. We analysed the spectrum of the reflected light from each

epithelium and determined the nuclear-size distribution of the epithelial cells.

We then used this nuclear-size distribution to obtain the percentage of nuclei larger than 10 μm across and the total number of nuclei per unit area (population density). As already noted, these parameters quantitatively characterize the degree of nuclear enlargement and crowding, respectively. Figure 1 shows these light-scattering spectroscopy parameters as binary plots to indicate the degree of correlation with histological diagnosis: in all four organs, there is a marked distinction between dysplastic and non-dysplastic epithelium. Both dysplasia and carcinoma *in situ* are associated with a higher percentage of enlarged nuclei and, on average, a higher population density, which can be used as the basis for spectroscopic diagnosis.

Our results show that light-scattering spectroscopy has the potential to detect epithelial precancerous lesions and

preinvasive cancers throughout the body. The advantage of this technique over conventional diagnostic procedures is that it can provide objective, quantitative results in real time without the need for tissue removal. Because of its potential application in screening for endoscopically invisible lesions, this technique should significantly improve the efficiency of cancer screening and surveillance.

V. Backman*, M. B. Wallace†, L. T. Perelman*, J. T. Arendt*, R. Gurjar*, M. G. Müller*, Q. Zhang*, G. Zonios*, E. Kline‡, T. McGillican§, S. Shapshay§, T. Valdez§, K. Badizadegan||, J. M. Crawford¶, M. Fitzmaurice#, S. Kabani§, H. S. Levin☆, M. Seiler**, R. R. Dasari*, I. Itzkan*, J. Van Dam†, M. S. Feld*

*Laser Biomedical Research Center, G. R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
e-mail: msfeld@MIT.edu

†Division of Gastroenterology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA

‡Department of Urology, Cleveland Clinic Foundation, Cleveland, Ohio, 44106, USA

§New England Medical Center, Boston, Massachusetts 02132, USA

||Department of Pathology, Children's Hospital, Boston, Massachusetts 02115, USA

¶Department of Pathology, Yale University School of Medicine, New Haven, Connecticut 06520-8023, USA

#Department of Pathology, University Hospitals of Cleveland and Case Western Reserve University, Cleveland, Ohio 44106, USA

☆Department of Pathology, Cleveland Clinic Foundation, Cleveland, Ohio 44106, USA

**US Veteran Administration Hospital, West Roxbury, Massachusetts 02132, USA

1. Cotran, R. S., Robbins, S. L. & Kumar, V. *Robbins Pathological Basis of Disease* (Saunders, Philadelphia, 1994).
2. Van de Hulst, H. C. *Light Scattering by Small Particles* (Dover, New York, 1957).
3. Perelman, L. T. et al. *Phys. Rev. Lett.* **80**, 627–630 (1998).
4. Backman, V. et al. *IEEE J. Sel. Top. Quant.* **5**, 1019–1026 (1999).

sequence tags of preparasitic juveniles (J2s) of the potato-cyst nematode *Globodera rostochiensis*. One expressed sequence tag (ge222) encoded a partial open reading frame with similarity to bacterial and fungal pectate lyases (EC 4.2.2.2). A full-length complementary DNA sequence, which included the ge222 tag, was subsequently obtained by the rapid amplification of cDNA ends by using messenger RNA from *G. rostochiensis* J2s as starting material².

The full-length cDNA contained a predicted open reading frame encoding a peptide sequence (PEL-1) of relative molecular mass 28K, with a signal sequence for secretion at its amino terminus. We expressed this coding region in *Pichia pastoris* to produce the active pectate lyase. We also found active pectate lyase in homogenates of *G. rostochiensis* J2. A digoxigenin-labelled DNA probe amplified from *pel-1* hybridized specifically to the subventral oesophageal gland cells, which also secrete cellulases³. These oesophageal gland cells are free of any symbiotic microorganisms³.

The PEL-1 sequence shows highest similarity (*E*-value of $5e^{-19}$ in BLAST-P⁴) with class III pectate lyases of bacterial and fungal origin (27–34% identity and 46–53% amino-acid similarity)⁴. Pectate lyases are divided into five classes, of which only class I (the pectate lyase superfamily) and class II (for example, periplasmic pectate lyase of *Erwinia carotovora*) have been characterized⁵; structural and functional information is sparse for class III pectate lyases.

Classification of PEL-1 as a class III pectate lyase markedly reduces the number of invariant amino-acid residues in this group (Fig. 1). Class III pectate lyases have four highly conserved regions and are characterized by the presence of several cysteine residues (eleven in PEL-1). On the basis of comparison of the PEL-1 sequence with other class III pectate lyases, only four conserved amino-acid residues with charged side chains (two aspartate residues at positions 130 and 199, and two lysine residues at positions 137 and 160) are present in the conserved regions. It is therefore likely that one or more of these invariant residues are involved in the catalytic machinery of PEL-1 and other class III members.

Symbiont-independent degradation of plant-cell walls by animals is now recognized as being possible. An endogenous cellulase gene was first isolated from cyst nematodes²; cellulases are also produced by termites⁶ and the redclaw crayfish (*Cherax quadricarinatus*)⁷. Our current finding demonstrates that, like bacteria and fungi, cyst nematodes are genetically equipped to secrete a mixture of depolymerizing cellulase and pectinase enzymes that allow the basic framework of plant cell walls to be dismantled.

Herman Popeijus*, Hein Overmars*, John Jones†, Vivian Blok†, Aska Goverse*,

Enzymology

Degradation of plant cell walls by a nematode

Interwoven networks of cellulose and pectin are the main components of plant cell walls¹, making them recalcitrant structures that can only be degraded by organisms producing a mix of synergistically acting enzymes. Animals were believed

to be unable to synthesize these enzymes, depending instead on symbiotic microbes to render plants into a food source. Here we describe a metazoan pectinase gene that encodes a pectate lyase for breaking down the pectin component of plant cell walls. To our knowledge, this is the first example of non-symbiotic degradation of pectin in plant cell walls by an animal.

We cloned the pectate lyase gene as part of a project analysing 1,000 expressed

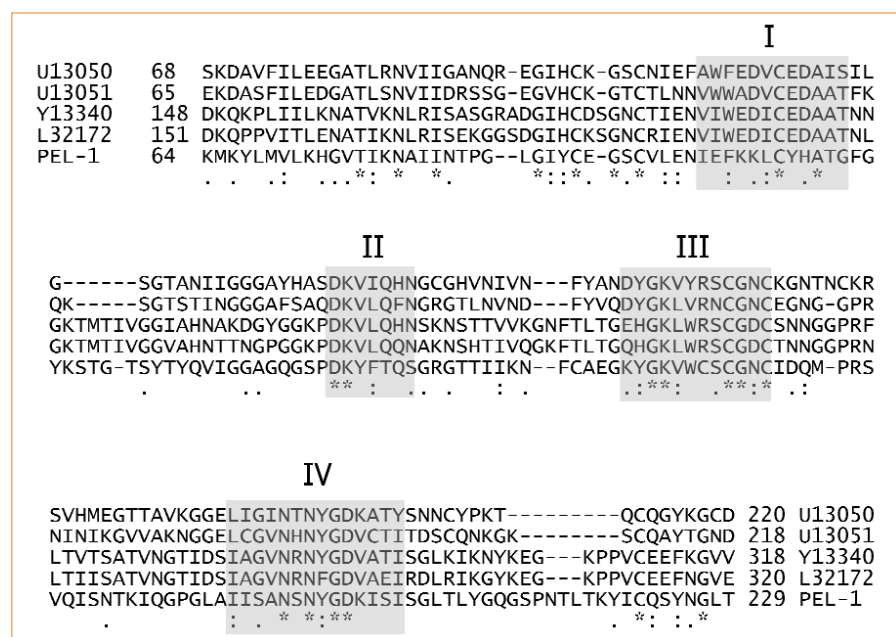


Figure 1 Predicted amino-acid sequence of *Globodera rostochiensis* PEL-1 (Genbank accession no. AF127915) aligned with bacterial (accession no. Y13340, *Erwinia chrysanthemi* Pel I, accession no. L32172, *E. carotovora* PelA) and fungal (accession no. U13050, *Nectria haematococca* Pel D, and accession no. U13051, *N. haematococca* Pel B) class III pectate lyases. Boxed shaded residues are conserved regions I to IV, as described for class III pectate lyase⁵. The alignment was made by using the program ClustalW and the BLOSUM62 substitution matrix. Asterisks, identical or conserved residues in all sequences in the alignment; colons, conserved substitutions; single dots, semiconserved substitutions⁸.

Johannes Helder*, Arjen Schots†, Jaap Bakker*, Geert Smant*

Laboratories of *Nematology and †Monoclonal Antibodies, Graduate School of Experimental Plant Science, Department of Plant Science, Wageningen University, Binnenhaven 10, 6708 PD Wageningen, The Netherlands

e-mail: geert.smant@medew.nema.wau.nl

†Nematology Department, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

1. Carpita, N. C. & Gibeault, D. M. *Plant J.* **3**, 1–30 (1993).
2. Smant, G. et al. *Proc. Natl Acad. Sci. USA* **95**, 4906–4911 (1998).
3. De Boer, J. M. et al. *Mol. Plant–Microbe Interact.* **9**, 39–46 (1996).
4. Altschul, S. F. et al. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
5. Shevchik, V. E., Robert-Baudouy, J. & Hugovieux-Cotte-Pattat, N. *J. Bacteriol.* **179**, 7321–7330 (1997).
6. Watanabe, H., Noda, H., Tokuda, G. & Lo, N. *Nature* **394**, 330–331 (1998).
7. Byrne, K. A., Lehnert, S. A., Johnson, S. E. & Moore, S. S. *Gene* **239**, 317–324 (1999).
8. Thompson, J. D., Higgins, D. G. & Gibson, T. J. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

Bacterial metabolism

Phosphite oxidation by sulphate reduction

Biological phosphorus occurs almost exclusively as phosphate in the redox state of +V, although a few phosphonic (+III) and phosphinic (+I) acids are found as secondary metabolites¹ or as constituents of phosphonolipids. Here we show that a culture of a lithoautotrophic bacterium purified from marine sediments in Venice can grow by anaerobic oxidation of phosphite (+III) to phosphate (+V) while simultaneously reducing sulphate to hydrogen sulphide. To our knowledge, this is the first description of a redox reaction involving phosphorus in microbial energy metabolism, an activity that might have operated on the early Earth and which could represent an ancient evolutionary trait.

The reduction of phosphate to phosphine (PH₃) has to proceed through steps of extremely low redox potential² (HPO₄^{2−}/HPO₃^{2−}, −690 mV; HPO₃^{2−}/H₂PO₂[−], −913 mV; H₂PO₂[−]/P, −922 mV; P/PH₃, −525 mV, at pH 7.0), so it is unlikely to drive the coupling of a respiratory process to biomass oxidation^{3,4}. Traces of phosphine have been detected in sediments⁵, paddy fields⁶ and manure samples⁴, but their origin is unclear.

Phosphite oxidation to phosphate should provide a good electron source for microbial energy metabolism, however, as the electrons are released at a very low redox potential. Phosphite and hypophosphite can serve as a phosphorus source for aerobic^{7,8} and anaerobic bacteria⁹, but so far there has been no evidence for quantitative phosphite oxidation as a type of energy metabolism.

We enriched phosphite-oxidizing, anaerobic bacteria in a sulphide-reduced mineral medium¹⁰ with sulphate as the elec-

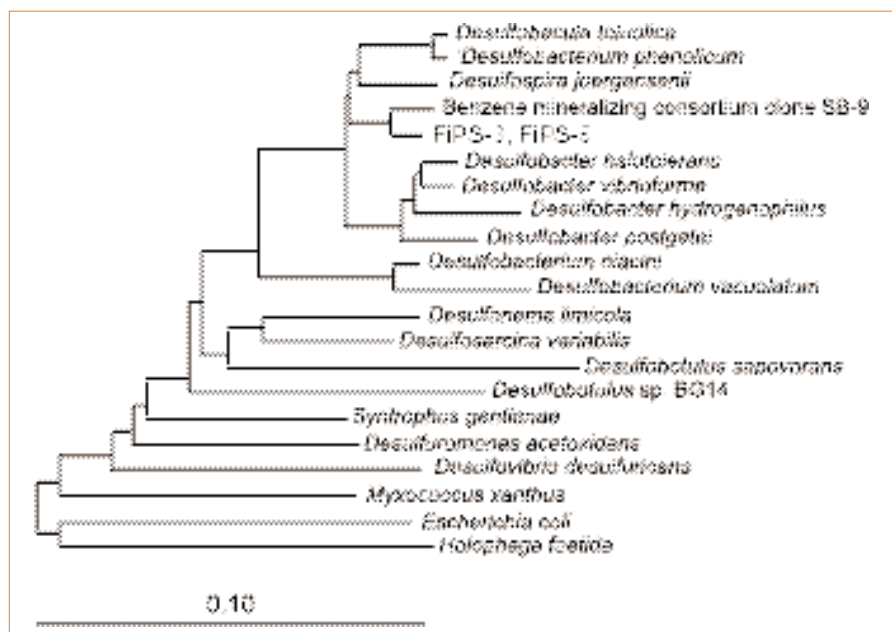


Figure 1 Phylogenetic tree based on gene sequences encoding the small subunit of ribosomal RNA (16S rRNA), showing the relation of the phosphite-oxidizing, sulphate-reducing strain FIPS-3 to members of the genus *Desulfobacter*, *Desulfobacula*, *Desulfospira*, and other representatives of major lineages of the δ -subclass of Proteobacteria. Scale bar represents 10% estimated difference in nucleotide sequences. DNA from strain FIPS-3 was extracted and purified using standard protocols and the rDNA was sequenced and analysed by using the ARB software package (TU Munich). The topology of the phylogenetic tree was derived from distance-matrix analyses and constructed using a neighbour-joining algorithm. *Escherichia coli* and *Holophaga foetida* were used as outgroup references.

tron acceptor, by using marine or freshwater sediments as inoculum. Growing cultures developed within 2–4 months and were transferred repeatedly into subcultures. Control untreated cultures, or cultures containing autoclaved cells, showed no evidence of phosphite oxidation. We isolated a pure culture of a strictly anaerobic, phosphite-oxidizing sulphate-reducing bacterium, strain FIPS-3, which grew with phosphite plus sulphate and a doubling time of three days according to

$4\text{HPO}_3^{2-} + \text{SO}_4^{2-} + \text{H}^+ \rightarrow 4\text{HPO}_4^{2-} + \text{HS}^-$
where $\Delta G' = -364$ kJ per mol sulphate, or -91 kJ per mol phosphite.

The strain also oxidized fumarate and malate to CO₂, and slowly oxidized hydrogen and formate. Lithotrophic growth with phosphite or hydrogen was not enhanced by addition of acetate, indicating that this strain covers all its carbon assimilation through autotrophic CO₂ fixation.

Analysis of 16S ribosomal RNA gene sequences revealed that strain FIPS-3 shares strong sequence similarities with sulphate-reducing Proteobacteria of the δ -subclass and is most closely related to an as-yet undescribed benzene-mineralizing clone SB-9 and to *Desulfospira joergensenii*, '*Desulfobacterium phenolicum*' and *Desulfobacula toluolica* (Fig. 1).

The ecological contribution of dissimilatory phosphite oxidation is unclear; on today's oxygen-exposed Earth, phosphite and other reduced inorganic phosphorus compounds are unstable. Phosphite may be an intermediate in the microbial degrada-

tion of organophosphonates in anoxic habitats, but there is no evidence for this. In prebiotic and archaean times, reduced phosphorus compounds might have been important as precursors of biochemical phosphorus compounds¹¹, for example, or they may have been introduced by meteorites¹². Anaerobically respiring microorganisms like the sulphate-reducing bacterium described here could have thrived on a litho(auto)trophic metabolism involving the oxidation of reduced phosphorus compounds, in which case dissimilatory phosphite oxidation by sulphate-reducing bacteria may represent an ancient evolutionary trait.

Bernhard Schink*, Michael Friedrich†

*Faculty of Biology, University of Konstanz, PO Box 5560, 78457 Konstanz, Germany

e-mail: Bernhard.Schink@uni-konstanz.de

†Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse, 35043 Marburg, Germany

1. Ternan, N. G., McGrath, J. W., McMullan, G. & Quinn, J. P. *World J. Microbiol. Biotechnol.* **14**, 635–647 (1998).
2. West, R. C., Astle, M. J. & Beyer, W. H. (eds) *CRC Handbook of Chemistry and Physics* (CRC, Boca Raton, FL, 1988).
3. Devai, L., Felföldy, L., Wittner, I. & Plosz, S. *Nature* **333**, 343–345 (1988).
4. Gassmann, G. & Glindemann, D. *Angew. Chem.* **32**, 761–763 (1993).
5. Gassmann, G. & Schorn, F. *Naturwissenschaften* **80**, 78–80 (1993).
6. Tsubota, G. *Soil Plant Food* **5**, 10–15 (1959).
7. Malacinski, G. & Konecny, W. A. *J. Bacteriol.* **91**, 578–582 (1966).
8. Metcalf, W. W. & Wolfe, R. S. *J. Bacteriol.* **180**, 5547–5558 (1998).
9. Foster, T. L., Winans, L. & Elms, S. J. *Appl. Environ. Microbiol.* **35**, 937–944 (1978).
10. Widdel, F. & Pfennig, N. *Arch. Microbiol.* **129**, 395–400 (1981).
11. Schwartz, A. W. *J. Theor. Biol.* **187**, 523–527 (1997).
12. Cooper, G. W., Onwo, W. M. & Cronin, J. R. *Geochim. Cosmochim. Acta* **56**, 4109–4115 (1992).

Inspiration for optimization from social insect behaviour

E. Bonabeau^{*†}, M. Dorigo[‡] & G. Theraulaz[§]

^{*} Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA

[†] Eurobios, Tour Ernst & Young, 92037 La Défense Cédex, France

[‡] IRIDIA, Université Libre de Bruxelles, CP 194/6, Avenue Franklin Roosevelt 50, 1050 Brussels, Belgium

[§] Laboratoire d'Ethologie et Cognition Animale, CNRS-FRE 2041, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cédex, France

Research in social insect behaviour has provided computer scientists with powerful methods for designing distributed control and optimization algorithms. These techniques are being applied successfully to a variety of scientific and engineering problems. In addition to achieving good performance on a wide spectrum of 'static' problems, such techniques tend to exhibit a high degree of flexibility and robustness in a dynamic environment.

Ethologists use modelling to understand animal behaviour. Recent research in social insect behaviour suggests that models based on self-organization can help explain how complex colony-level behaviour emerges out of interactions among individual insects^{1–3}. Although the goal of a modeller is generally to understand the living, a model can in principle be explored beyond the biologically plausible. Although biology does not necessarily benefit from such an exploration, computer scientists and engineers have been able to transform models of social insect collective behaviour into useful optimization and control algorithms. This new line of research concerns the transformation of knowledge about how social insects collectively solve problems into artificial problem-solving techniques—producing a form of Artificial Intelligence, or Swarm Intelligence³, in which the underlying model of intelligence is the collective intelligence of a social insect colony. Some of the techniques of Swarm Intelligence are now maturing. Optimization and control algorithms inspired by models of co-operative food retrieval in ants have been unexpectedly successful and have become known in recent years as Ant Colony Optimization (ACO)^{4,5} and Ant Colony Routing (ACR)^{6–8}. Real-world implementations are emerging. Other techniques, inspired by co-operative brood sorting by ants^{9,10} or task allocation in social wasps³, are still in a preliminary, proof-of-concept stage, with no systematic benchmarking of their performance. ACO and ACR will be described in more detail here.

Ant Colony Optimization

The ability of ants to optimize is best exemplified by the experiment described in Box 1. In this experiment¹, the ants can select the shortest path to a food source because they lay and follow pheromone (chemical) trails. But the ants' success in collectively selecting the shortest path is only statistical: the colony may occasionally get 'stuck' on a longer path if by chance the longer path is the first one to be marked. In using the 'trail laying–trail following' metaphor for optimization purposes, computer scientists have found it essential to improve the convergence properties of their algorithms by artificially increasing the rate of pheromone evaporation beyond biological plausibility.

The classic Travelling Salesman Problem (TSP), which consists of finding the shortest tour between n cities visiting each once only and ending at the starting point, illustrates how the use of artificial pheromones can be applied to hard optimization problems^{4,5}. Each artificial ant visits the n cities to construct a tour. After completing its tour it lays 'artificial pheromone' on the links it has used in an amount that is proportional to the quality of the tour, so that links which belong to good solutions end up with more pheromone than

the other links. To construct a tour, each ant tends to select a city connected to its current city by a link that has more pheromone than the others: this selection process amplifies previously reinforced links and leads to the emergence of a good solution. Pheromone evaporates at a fixed rate after all ants have constructed their tours. Evaporation prevents mediocre links from being amplified by accident. The mathematical description of the simplest ACO algorithm described in Box 2 is a proof-of-concept, which performs poorly on large TSP instances. More sophisticated versions include a local search that seeks to improve existing solutions by evaluating the effects of small changes (such as swapping two cities), a more complex tradeoff between exploration and exploitation, increased reinforcement of the best tour, and other refinements. Such improvements make the algorithm much more competitive, in terms of both computing time and quality of solutions^{5,40}: comparisons of an improved ACO algorithm with other efficient, state-of-the-art implementations of general-purpose approaches for the TSP

Box 1

How ants find the shortest path

Ant colonies can collectively perform tasks and make decisions that appear to require a high degree of co-ordination among the workers: building a nest, feeding the brood, foraging for food, and so on. In the example presented here, the ants collectively discover the shortest path to a food source. In experiments with the ant *Linepithema humile*, a food source is separated from the nest by a bridge with two branches^{1,2}. The longer branch is r times longer than the shorter branch (Fig. 2a).

The shorter branch is selected by the colony in most experiments if r is sufficiently large ($r = 2$ in Fig. 2b). This is because the ants lay and follow pheromone trails: individual ants lay a chemical substance, a pheromone, which attracts other ants. The first ants returning to the nest from the food source are those who take the shorter path twice (from the nest to the source and back). At choice points 1 and 2, nestmates are recruited toward the shorter branch, which is the first to be marked with pheromone. If, however, the shorter branch is presented to the colony after the longer branch, the shorter path will not be selected because the longer branch is already marked with pheromone (Fig. 2b).

This problem can be overcome in an artificial system, by introducing pheromone decay: when the pheromone evaporates sufficiently quickly, it is more difficult to maintain a stable pheromone trail on a longer path. The shorter branch can then be selected even if presented after the longer branch. This property is particularly desirable in optimization, where one seeks to avoid convergence toward mediocre solutions. In *Linepithema humile*, although pheromone concentrations do decay, the lifetime of trail pheromones is too large to allow such flexibility.

Box 2

Ant Colony Optimization and the Travelling Salesman Problem

The Travelling Salesman Problem (TSP) consists of finding the shortest tour between n cities visiting each once only and ending at the starting point. Let d_{ij} be the distance between cities i and j and τ_{ij} the amount of pheromone on the edge that connects i and j . τ_{ij} is initially set to a small value τ_0 , the same for all edges (i, j) . The algorithm consists of a series of iterations. One iteration of the simplest ACO algorithm applied to the TSP can be summarized as follows: (1) a set of m artificial ants are initially located at randomly selected cities; (2) each ant, denoted by k , constructs a complete tour, visiting each city exactly once, always maintaining a list J_k of cities that remain to be visited; (3) an ant located at city i hops to a city j , selected among the cities that have not yet been visited, according to probability $p_{ij}^k = (\tau_{ij})^\alpha \cdot (d_{ij})^{-\beta} / (\sum_{l \in J_k} (\tau_{il})^\alpha \cdot (d_{il})^{-\beta})$, where α and β are two positive parameters which govern the respective influences of pheromone and distance; (4) when every ant has completed a tour, pheromone trails are updated: $\tau_{ij} \leftarrow (1 - \rho) \tau_{ij} + \Delta \tau_{ij}$, where ρ is the evaporation rate and $\Delta \tau_{ij}$ is the amount of reinforcement received by edge (i, j) . $\Delta \tau_{ij}$ is proportional to the quality of the solutions in which (i, j) was used by one ant or more. More precisely, if L_k is the length of the tour T_k constructed by ant k , then $\Delta \tau_{ij} = \sum_{k=1}^m \Delta \tau_{ij}^k$, with $\Delta \tau_{ij}^k = Q/L_k$ if $(i, j) \in T_k$ and $\Delta \tau_{ij}^k = 0$ otherwise, where Q is a positive parameter. This reinforcement procedure reflects the idea that pheromone density should be lower on a longer path because a longer trail is more difficult to maintain.

Steps (1) to (4) are repeated either a predefined number of times or until a satisfactory solution has been found. The algorithm works by reinforcing portions of solutions that belong to good solutions and by applying a dissipation mechanism, pheromone evaporation, which ensures that the system does not converge early toward a poor solution. When $\alpha = 0$, the algorithm implements a probabilistic greedy search, whereby the next city is selected solely on the basis of its distance from the current city. When $\beta = 0$, only the pheromone is used to guide the search, which would reflect the way the ants do it. However, the explicit use of distance as a criterion for path selection appears to improve the algorithm's performance^{4,5}. In all other optimization applications also, an improvement in the algorithm's performance is observed when a local measure of greed, similar to the inverse of distance for the TSP, is included into the local selection of portions of solution by the agents. Typical parameter values are: $m = n$, $\alpha = 1$, $\beta = 5$, $\rho = 0.5$, $\tau_0 = 10^{-6}$.

(genetic algorithm²⁷, iterated local search and iterated Lin-Kernighan, see <http://www.keck.caam.rice.edu/concorde.html>) show that the ACO algorithm is competitive, always providing among the best results.

In addition to finding very good solutions to 'static' problems, the algorithm also maintains a pool of alternative portions of solutions, which can be especially useful when the problem is dynamically changing: the algorithm can channel new searches toward this pool. This is because the pheromone update equation ensures that every link has at least a small amount of pheromone: a weakly used link can be quickly reinforced and replace a missing or failing link. Although there is no formal framework for quantifying the performance of optimization algorithms in dynamic environments, the flexibility and robustness of an algorithm are important factors when it comes to real-world implementations.

The TSP is a natural application of ACO, but many other combinatorial optimization problems can be solved with ACO. ACO is currently the best available heuristic for the sequential ordering problem and for real-world instances of the quadratic assignment problem, and is among the most competitive approaches for the vehicle and network routing problems, as well as for a number of other problems (see Table 1). Regarding real-world applications, an ACO implementation for multi-stage factory scheduling is under study at Unilever (D. Gregg *et al.*, personal communication), and gasoline trucks are now being routed with the help of an ACO algorithm in Italian Switzerland (L. M. Gambardella, personal communication).

ACO, however, does not always work well. For example, it does not perform as well as other heuristics on any problem's instances that have been uniformly randomly generated³. The reason is that ACO tends to reinforce portions of solutions that belong to many good solutions: the more good solutions a given portion belongs to, the more virtual pheromone it receives. If a large number of portions of solutions are equally likely to be part of good solutions, ACO cannot differentiate them and therefore performs poorly. However, many real-world problems do possess enough of the requisite 'structure' to allow this approach to be very efficient. Other optimization techniques, based on exotic neural networks¹¹ or evolutionary algorithms¹², rely on the same 'building block' principle, that is, they attempt to find the building blocks of good solutions by maintaining a search statistics. In ACO, the search statistics is summarized into pheromone concentrations, which, after a sufficient number of iterations, indicate the likelihood that particular portions of solution, or building blocks, belong to one of the best solutions. ACO not only identifies building blocks but also identifies correlations between building blocks: for example, two or more portions of solution may be desirable only if used together. This property is crucial, as most instances of most problems are not readily 'linearly' decomposable into building blocks.

Ant Colony Routing in communications networks

ACO has an additional feature that makes it particularly appealing: it is explicitly formulated in terms of computational agents. Although it may be, in principle, possible to focus only on the core optimizing mechanism (reinforcement of portions of solutions and global dissipation), the agent-based formulation is a useful aid for designing problem-solving systems. The example of routing in communications networks illustrates this point.

Routing is the control mechanism that directs every message in a communications network from its source node to its destination node through a sequence of intermediate nodes or switching stations. Each switching station has a routing table that tells messages or portions of messages called packets where to go, given their destinations. Because of the highly dynamic nature of communications networks, due to the time-varying stochastic changes in user traffic patterns as well as to unpredictable failures of network components, portions of a network may become congested and new routes have to be discovered dynamically. In Ant Colony Routing (ACR), first described by Schoonderwoerd *et al.*⁶, ant-like agents reinforce routing-table entries depending on their experience and performance in the network⁶⁻⁸. For example, if

Table 1 Some applications of Ant Colony Organization and Ant Colony Routing

Problem name	Year	Main references	Algorithm name
Travelling salesman	1991	4	AS
	1997	5	ACS & ACS-3-opt
	1997	28	MMAS
Network routing	1997	6	ABC
	1998	7	Co-operative Asymmetric Forward AntNet
	1998	8	AntNet
	1999	29	ABC-backward
Graph colouring	1997	30	ANTCOL
Shortest common supersequence	1999	31	AS-SCS
Quadratic assignment	1999	32	HAS-QAP
	1999	33	MMAS-QAP
	1999	34	AS-QAP
Machine scheduling	1999	35	ACS-SMTTP
Vehicle routing	1999	36	AS-VRP
	1999	37	MACS-VRPTW
Multiple knapsack	1999	38	AS-MKP
Frequency assignment	2000	39	ANTS-FAP
Sequential ordering	1997	40	HAS-SOP

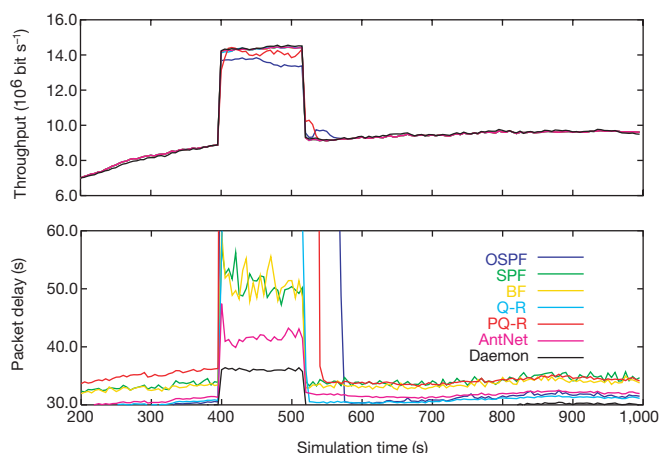


Figure 1 Typical result of a comparison of AntNet, an Ant Colony Routing algorithm, with other widespread routing algorithms for packet-switched networks (see ref. 41 for an overview of communications networks). OSPF is the Open Shortest Path First algorithm, the current official Internet routing algorithm^{3,8}. BF is the asynchronous distributed Bellman-Ford algorithm with a dynamic cost metric^{3,8,41}. SPF is the Shortest Path First algorithm with a dynamic cost metric^{3,8}. Q-R is the Q-Routing algorithm^{3,8}. PQ-R is the Predictive Q-Routing algorithm^{3,8}, an extension of Q-Routing. Daemon is an approximation of an ideal algorithm and provides a bound to the best performance achievable^{3,8}. The network, the underlying architecture of which is that of NSFnet, is highly loaded with stochastic time-varying traffic. At time 400 a sudden increase in traffic, lasting 120 s, takes the network to saturation conditions. The upper graph compares throughput (the larger the better), while the lower graph packet delays averaged over a 5-s time window (the smaller the better). AntNet provides the same throughput as the best competing algorithms maintaining a much lower average packet delay. After ref. 8, courtesy of the AI Access Foundation and Morgan Kaufmann.

an agent has been delayed a long time because it went through a highly congested portion of the network, it will weakly reinforce routing-table entries that send messages to that portion of the network; an agent that enjoyed fluid traffic conditions will apply a stronger reinforcement. A dissipation or evaporation mechanism is also applied regularly to the routing-table entries to refresh the system and prevent obsolete solutions from being maintained.

When tested on simulated networks and realistic traffic conditions, ACR appears to outperform routing algorithms which are in widespread use, especially, but not only, in strongly variable traffic conditions⁸ (Fig. 1). Maintaining a pool of alternate routes is the way the system copes with changing conditions, including node failure, and this technique makes it flexible and robust. Although some of the reported results are quite good, intensive real-world testing would be useful. Recent news articles indicate that British

Telecom and MCI-Worldcom have been and still are using algorithms based on the ant colony metaphor¹³.

Other applications inspired by social insects

Ant foraging is not the only social insect behaviour that has inspired computer scientists and roboticists. Other examples include division of labour, brood sorting and co-operative transport.

Division of labour. In most social insect species, individual workers tend to be specialized in certain tasks for a varying portion of their lifetime¹⁴. But the behavioural repertoire of workers can be stretched back and forth in response to perturbations: if needed, a forager can become a nurse, or a nurse a guard, and so forth. Such a combination of specialization and plasticity in task allocation is appealing for multi-agent optimization and control. A response threshold model of division of labour has been used to solve dynamic task scheduling problems^{3,15}. In a threshold model¹⁴, workers with low response thresholds respond to lower levels of stimuli than do workers with high response thresholds. Task performance reduces the intensity of stimuli. If workers with low thresholds perform their normal tasks the task-associated stimuli never reach the thresholds of the high-threshold workers. But if, for any reason, the intensity of task-associated stimuli increases, high-threshold workers engage in task performance. For example, to allocate trucks coming out of an assembly line to paint booths in a truck factory, each paint booth is considered to be an insect-like agent that is specialized in one colour; but if needed, the paint booth can change its colour (although this is costly). Thus, the swarm intelligent system minimizes task changeovers and can cope with glitches³.

Brood sorting. In the ant *Leptothorax unifasciatus*, workers sort the brood¹⁶. Eggs and microlarvae are compactly clustered at the centre of the brood area; the largest larvae are located at the periphery of the brood cluster; when pupae and prepupae are present, they are located between peripheral large larvae and the more central larvae of medium size. Deneubourg *et al.*¹⁷ have proposed a model of this phenomenon, recognizing that many of their model's assumptions had to be tested. In the model, an ant picks up and drops an item according to the number of similar surrounding items. For example, if an ant carries a large larva, it will drop the larva with high probability in a region populated by large larvae, or if an unladen ant finds a large larva surrounded by eggs, it will pick up the larva with high probability; in all other situations the ant will neither drop nor pick up any item. Although the model still needs validation, computer scientists have found it useful for data sorting. Lumer and Faieta⁹ and Kuntz *et al.*¹⁰ have applied it to the following problem. Given a data set of points in an n -dimensional space and a metric δ which measures the distance between pairs of data points, project the points onto the plane so that if any two projected points in the plane are neighbours, their corresponding data items are

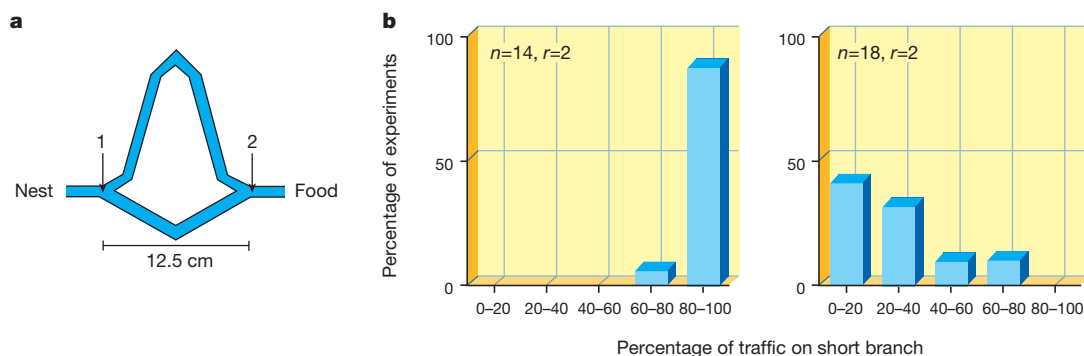


Figure 2 The double-bridge experiment. **a**, Experimental set-up. **b**, Distribution of the percentage of ants that selected the shorter branch over a set of experiments. The longer branch is r times longer than the short branch. The left graph (14 experiments) represents experiments in which both branches were presented simultaneously. The right graph (18

experiments) represents experiments in which the shorter branch was presented to the colony 30 minutes after the longer branch: the shorter branch is not selected and the colony keeps on exploiting the longer branch. Modified from ref. 2.

neighbours under the metric δ . The initial projection of data items onto the plane is random. The artificial ants then perform random walks on the plane and pick up or drop projected data items using rules from the model¹⁷. The results are qualitatively equivalent to those obtained by such classic techniques as spectral decomposition or stress minimization¹⁰, but at a much lower computational cost.

Co-operative transport. Swarm Intelligence is also an inspiration for roboticists to design distributed control algorithms for groups, or swarms, of robots^{16–22}. One example of a task that has been used as a benchmark for swarm robotics is co-operative transport, or more precisely co-operative box pushing^{20,21}. Co-operative transport has been reported in several ant species^{23–25}. When it is impossible for a single ant to retrieve a large prey item, nestmates are recruited. Then, for a few minutes, the ants change position and alignment around the prey item without making progress, until it can be moved toward the nest. This emergent co-ordination has been reproduced by Kube and Zhang²¹ in a swarm of very simple robots whose task is to push a box toward a goal. Their work shows how a task that appears to require the co-ordinated efforts of several robots can be performed without explicit co-ordination or communication among the robots. Such work is promising in the perspective of miniaturization and low-cost robotics (see for example the Alice micro-robots at http://dmtwww.epfl.ch/isr/asl/projects/alice_pj.html).

Perspectives

The initial appeal of Swarm Intelligence to computer scientists was almost entirely due to their fascination with ants. The surprisingly good results obtained by several groups of researchers make it even more appealing for optimization and control applications, especially in view of the ability of swarm intelligent systems to cope with glitches and changing environments.

A number of problems, best exemplified by routing in telecommunications networks, are simply much better tackled with computational agents. The breakthrough brought by Swarm Intelligence is permitting the design of agent-based, distributed optimization and control techniques. We view Swarm Intelligence as a major new paradigm in optimization and control. Its closest relative would be the artificial neural network paradigm. Indeed, an ant colony is a 'connectionist' system, that is, one in which individual units are connected to each other according to a certain pattern. However, crucial differences from typical neural networks include: (1) the dynamic nature of the connectivity pattern, which can be exploited by examining how ants or other social insects solve problems; (2) the explicit or implicit mobility of the units, which is an essential feature of ACO and ACR; (3) feedback from the environment, which is used as a medium of co-ordination and communication; and (4) the use of pheromone evaporation in ACO, which dramatically facilitates the design of distributed optimization techniques.

We have no doubt that more practical applications of Swarm Intelligence will continue to emerge. In a world where a chip will soon be embedded into every object, from envelopes to trash cans to heads of lettuce, control algorithms will have to be invented to let all these 'dumb' pieces of silicon communicate²⁶. Meanwhile, a better understanding of how and why methods based on social insects work and more systematic comparisons with other heuristics for optimization and control are needed. □

1. Deneubourg, J. -L. & Goss, S. Collective patterns and decision making. *Ethol. Ecol. Evol.* **1**, 295–311 (1989).
2. Goss, S., Aron, S., Deneubourg, J. -L. & Pasteels, J. M. Self-organized shortcuts in the Argentine ant. *Naturwissenschaften* **76**, 579–581 (1989).
3. Bonabeau, E., Dorigo, M. & Theraulaz, G. *Swarm Intelligence: From Natural to Artificial Systems* (Oxford Univ. Press, New York, 1999).
4. Dorigo, M., Maniezzo, V. & Colnari, A. The ant system: optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man Cybern. B* **6**, 29–41 (1996).

5. Dorigo, M. & Gambardella, L. M. Ant colonies for the traveling salesman problem. *BioSystems* **43**, 73–81 (1997).
6. Schoonderwoerd, R., Holland, O., Bruten, J. & Rothkrantz, L. Ant-based load balancing in telecommunications networks. *Adapt. Behav.* **5**, 169–207 (1997).
7. Heusse, M., Guérin, S., Snyers, D. & Kuntz, P. Adaptive agent-driven routing and load balancing in communication networks. *Adv. Compl. Syst.* **1**, 237–254 (1998).
8. Di Caro, G. & Dorigo, M. AntNet: Distributed stigmergetic control for communications networks. *J. Artif. Intell. Res.* **9**, 317–365 (1998).
9. Lumer, E. & Faieta, B. in *Proc. 3rd Intl Conf. Simulation of Adaptive Behavior: From Animals to Animats 3* (eds Cliff, D., Husbands, P., Meyer, J. -A. & Wilson, S. W.) 501–508 (MIT Press, Cambridge, MA, 1994).
10. Kuntz, P., Snyers, D. & Layzell, P. A stochastic heuristic for visualizing graph clusters in a bi-dimensional space prior to partitioning. *J. Heuristics* **5**, 327–351 (1999).
11. Chen, K. A simple learning algorithm for the traveling salesman problem. *Phys. Rev. E* **55**, 7809–7812 (1997).
12. Baluja, S. & Caruana, R. in *Proc. 12th Intl Conf. Machine Learning* (eds Prieditis, A. & Russell, S.) 38–46 (Morgan Kaufmann, Palo Alto, 1995).
13. Ward, M. There's an ant in my phone. *New Sci.* **2118**, 32–35 (1998).
14. Robinson, G. E. Regulation of division of labour in insect societies. *Annu. Rev. Entomol.* **37**, 637–665 (1992).
15. Sobkowski, A. et al. in *Bio-Computation and Emergent Computing* (eds Lundh, D., Olsson B. & Narayanan, A.) 36–45 (World Scientific, Singapore, 1997).
16. Franks, N. R. & Sendova-Franks, A. B. Brood sorting by ants: distributing the workload over the work surface. *Behav. Ecol. Sociobiol.* **30**, 109–123 (1992).
17. Deneubourg, J. -L. et al. in *Proc. 1st Conf. Simulation of Adaptive Behavior: From Animals to Animats* (eds Meyer, J. A. & Wilson, S. W.) 356–365 (MIT Press, Cambridge, MA, 1991).
18. Cao, Y. U., Fukunaga, A. S. & Kahng, A. B. Cooperative mobile robotics: antecedents and directions. *Autonomous Robots* **4**, 7–27 (1997).
19. Martinoli, A., Yamamoto, M. & Mondada, F. in: *Proc. 4th European Conf. Artificial Life* (eds Husbands, P. & Harvey, I.) (MIT Press, Cambridge, MA, 1997).
20. Nilsson, M. & Simsarian, K. T. in *Proc. 1995 IEEE/RSJ Intl Conf. Intelligent Robots and Systems* 556–561 (IEEE Computer Society Press, Los Alamitos, 1995).
21. Kube, C. R. & Zhang, H. Collective robotics: from social insects to robots. *Adaptive Behavior* **2**, 189–218 (1994).
22. Kube, C. R. & Zhang, H. Task modelling in collective robotics. *Autonomous Robots* **4**, 53–72 (1997).
23. Sudd, J. H. The transport of prey by ants. *Behaviour* **25**, 234–271 (1965).
24. Franks, N. R. Teams in social insects: group retrieval of prey by army ants (*Eciton burchelli*, Hymenoptera: Formicidae). *Behav. Ecol. Sociobiol.* **18**, 425–429 (1986).
25. Moffett, M. W. Cooperative food transport by an asiatic ant. *Nat. Geogr. Res.* **4**, 386–394 (1988).
26. Kelly, K. *New Rules for the New Economy* (Viking, New York, 1998).
27. Walters, T. in *Proc. PPSN V, Conference on Parallel Problem-Solving from Nature* (eds Eiben, A. E., Bäck, T., Schoenauer, M. & Schwefel, H.-S.) 813–822 (Springer, Berlin, 1998).
28. Stützle, T. & Hoos, H. in *Proc. IEEE Intl Conf. Evolutionary Computation* (eds Bäck, T., Michalewicz, Z. & Yao, X.) 309–314 (IEEE Computer Society Press, Los Alamitos, 1997).
29. Van der Put, R. & Rothkrantz, L. in *Simulation Practice and Theory* (in the press).
30. Costa, D. & Hertz, A. Ants can colour graphs. *J. Op. Res. Soc.* **48**, 295–305 (1997).
31. Michel, R. & Middendorf, M. in *Proc. PPSN V, Conference on Parallel Problem-Solving from Nature* (eds Eiben, A. E., Bäck, T., Schoenauer, M. & Schwefel, H.-S.) 692–701 (Springer, Berlin, 1998).
32. Gambardella, L. M., Taillard, E. D. & Dorigo, M. Ant colonies for the quadratic assignment problem. *J. Op. Res. Soc.* **50**, 167–176 (1999).
33. Stützle, T. & Hoos, H. in *Meta-Heuristics: Advances and Trends in Local Search Paradigms for Optimization* (eds Voss, S., Martello, S., Osman, I. H. & Roucairol, C.) 313–329 (Kluwer Academic, Boston, 1999).
34. Maniezzo, V. & Colnari, A. The ant system applied to the quadratic assignment problem. *IEEE Trans. Knowledge Data Engin.* **11**, 769–778 (1999).
35. Bauer, A., Bullnheimer, B., Hartl, R. F. & Strauss, C. in *Proc. Congr. Evolutionary Computation (CEC'99)* 1445–1450 (IEEE Press, Piscataway, NJ, 1999).
36. Bullnheimer, B., Hartl, R. F. & Strauss, C. in *Meta-Heuristics: Advances and Trends in Local Search Paradigms for Optimization* (eds Voss, S., Martello, S., Osman, I. H. & Roucairol, C.) 109–120 (Kluwer Academic, Boston, 1999).
37. Gambardella, L. M., Taillard, E. D. & Agazzi, G. in *New Ideas in Optimization* (eds Corne, D., Dorigo, M. & Glover, F.) 63–76 (McGraw-Hill, London, 1999).
38. Leguizamón, G. & Michalewicz, Z. in *Proc. Congr. Evolutionary Computation (CEC'99)* 1459–1464 (IEEE Press, Piscataway, NJ, 1999).
39. Maniezzo, V. & Carbonaro, A. An ANTS heuristic for the frequency assignment problem. *Future Generation Comput. Syst. J.* **16**, 927–935 (2000).
40. Gambardella, L. M. & Dorigo, M. Ant colony system hybridized with a new local search for the sequential ordering problem. *INFORMS J. Comput.* (in the press).
41. Bertsekas, D. & Gallager, R. *Data Networks* (Prentice Hall, Englewood Cliffs, 1992).

Acknowledgements

E. B. is supported by the Interval Research Fellowship at the Santa Fe Institute. E. B. and G. T. are supported in part by a grant from the GIS (Groupeement d'Intérêt Scientifique) Sciences de la Cognition. G. T. is supported by a grant from the Conseil Régional Midi-Pyrénées. M. D. acknowledges support from the Belgian FNRS, of which he is a Research Associate.

Correspondence and requests for materials should be addressed to E.B. (e-mail: eric.bonabeau@eurobios.com). Information and data on comparisons with other algorithms, including source codes, can be obtained from M.D. (e-mail: mdorigo@ulb.ac.be).

Quantum superposition of distinct macroscopic states

Jonathan R. Friedman, Vijay Patel, W. Chen, S. K. Tolpygo & J. E. Lukens

Department of Physics and Astronomy, The State University of New York, Stony Brook, New York 11794-3800, USA

In 1935, Schrödinger¹ attempted to demonstrate the limitations of quantum mechanics using a thought experiment in which a cat is put in a quantum superposition of alive and dead states. The idea remained an academic curiosity until the 1980s when it was proposed^{2–4} that, under suitable conditions, a macroscopic object with many microscopic degrees of freedom could behave quantum mechanically, provided that it was sufficiently decoupled from its environment. Although much progress has been made in demonstrating the macroscopic quantum behaviour of various systems such as superconductors^{5–9}, nanoscale magnets^{10–12}, laser-cooled trapped ions¹³, photons in a microwave cavity¹⁴ and C₆₀ molecules¹⁵, there has been no experimental demonstration of a quantum superposition of truly macroscopically distinct states. Here we present experimental evidence that a superconducting quantum interference device (SQUID) can be put into a superposition of two magnetic-flux states: one corresponding to a few microamperes of current flowing clockwise, the other corresponding to the same amount of current flowing anticlockwise.

The simplest SQUID (the radio frequency (r.f.) SQUID) is a superconducting loop of inductance L broken by a Josephson tunnel junction with capacitance C and critical current I_c . In equilibrium, a dissipationless supercurrent can flow around this loop, driven by the difference between the flux Φ that threads the loops and the external flux Φ_x applied to the loop. The dynamics of the SQUID can be described in terms of the variable Φ and are analogous to those of a particle of 'mass' C (and kinetic energy $\frac{1}{2}C\dot{\Phi}^2$) moving in a one-dimensional potential (Fig. 1a) given by the sum of the magnetic energy of the loop and the Josephson coupling energy of the junction:

$$U = U_0 \left[\frac{1}{2} \left(\frac{2\pi(\Phi - \Phi_x)}{\Phi_0} \right)^2 - \beta_L \cos(2\pi\Phi/\Phi_0) \right] \quad (1)$$

where Φ_0 is the flux quantum, $U_0 \equiv \Phi_0^2/4\pi^2L$ and $\beta_L \equiv 2\pi LI_c/\Phi_0$. For the parameters used in our experiment, this is a double-well potential separated by a barrier with a height depending on I_c . When $\Phi_x = \Phi_0/2$ the potential is symmetric. Any change in Φ_x then tilts the potential, as shown in Fig. 1a.

As required by quantum mechanics, in the dissipationless state the phase of the macroscopic superconducting wavefunction must vary continuously around the loop, increasing by an integer f times 2π when one winds once around the loop. The quantum number f defines the 'fluxoid' state of the SQUID. In Fig. 1a, the left (right) well corresponds to $f = 0$ (1) in which a static current ($>1 \mu\text{A}$ for our system) flows around the loop in such a way as to tend to cancel (augment) Φ_x . Classically, a transition between the $f = 0$ and $f = 1$ wells involves passage over the top of the barrier. Quantum mechanically, however, the system can tunnel through the barrier. For weak damping, the system has quantized energy levels that, considerably below the barrier, are localized in each well. At various values of Φ_x , levels in opposite wells will align, giving rise to resonant tunnelling between the wells⁵. During this interwell transition Φ changes by some fraction of Φ_0 , or—stated in different terms—the magnetic moment of the system changes by a macroscopic amount, in this case by over $10^{10} \mu_B$. Until now, however, there has been no evidence that the tunnelling process between these macroscopically distinct states could be coherent, that is, that the

SQUID could be put into a coherent superposition of two flux states in different wells.

Such a superposition would manifest itself in an anticrossing, as illustrated in Fig. 1b, where the energy-level diagram of two levels of different fluxoid states (labelled $|0\rangle$ and $|1\rangle$) is shown in the neighbourhood in which they would become degenerate without coherent interaction (dashed lines). Coherent tunnelling lifts the degeneracy (solid lines) so that at the degeneracy point the energy eigenstates are $1/\sqrt{2}(|0\rangle + |1\rangle)$ and $1/\sqrt{2}(|0\rangle - |1\rangle)$, the symmetric and anti-symmetric superpositions. The energy difference ΔE between the two states is given approximately by $\Delta E = \sqrt{\epsilon^2 + \Delta^2}$, where Δ is known as the tunnel splitting. The goal of the present work is to demonstrate the existence of such a splitting and, thereby the coherent superposition of macroscopically distinct flux states.

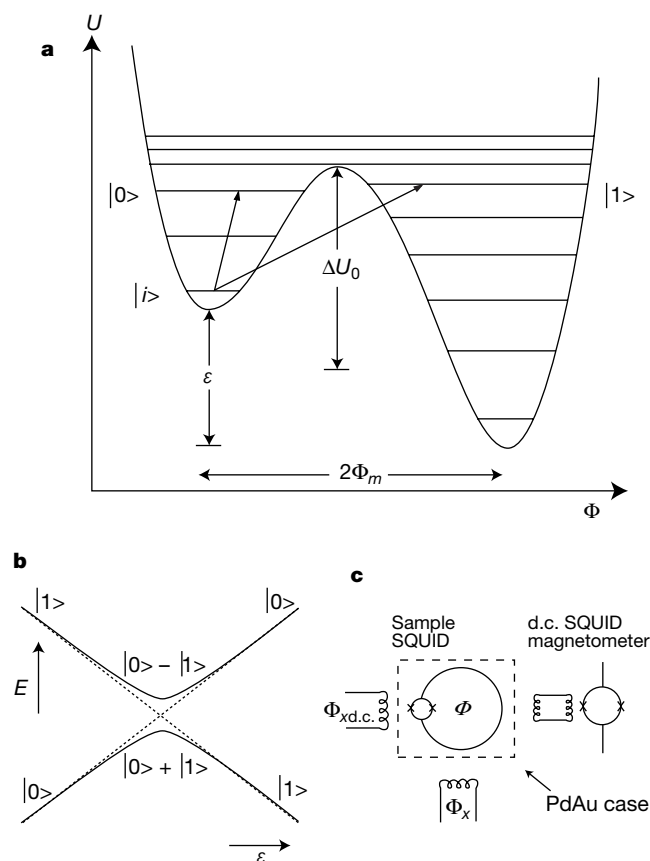


Figure 1 SQUID potential, energy-level anticrossing and experimental set-up. **a**, SQUID potential. The left well corresponds to the zero-fluxoid state of the SQUID and the right well to the one-fluxoid state. Energy levels are localized in each well. Both the tilt ϵ and energy barrier at zero tilt ΔU_0 can be varied *in situ* in the experiments. The process of photon-induced interwell transitions is illustrated by the arrows, where the system is excited out of the initial state $|i\rangle$ and into one of two excited states $|0\rangle$ or $|1\rangle$. **b**, Schematic anticrossing. When the two states $|0\rangle$ and $|1\rangle$ would become degenerate in the absence of coherence, the degeneracy is lifted and the states of the system are the symmetric and antisymmetric superpositions of the flux-basis states: $1/\sqrt{2}(|0\rangle + |1\rangle)$ and $1/\sqrt{2}(|0\rangle - |1\rangle)$. **c**, Experimental set-up. Our SQUID contains a 'tunable junction', a small d.c.-SQUID. A flux $\Phi_{x,d.c.}$ applied to this small loop tunes the barrier height ΔU_0 . Another flux Φ_x tunes the tilt ϵ of the potential. A separate d.c.-SQUID acts as a magnetometer, measuring the flux state of the sample. The sample SQUID used in our experiments is characterized by the following three energies: the charging energy $E_c \equiv e^2/2C = 9.0$ mK, the inductive energy $E_L \equiv \Phi_0^2/2L = 645$ K and a tunable Josephson coupling energy $E_J \equiv (I_c \Phi_0/2\pi) \cos(\pi \Phi_{x,d.c.}/\Phi_0) = 76 \text{ K} \cos(\pi \Phi_{x,d.c.}/\Phi_0)$. The plasma frequency ω_J (the frequency of small oscillations in the bottom of a well) associated with these parameters is $1.5\text{--}1.8 \times 10^{11} \text{ rad s}^{-1}$, depending on the value of $\Phi_{x,d.c.}$. The fact that $E_c \ll E_L, E_J$ confirms that flux is the proper basis to describe the SQUID's dynamics.

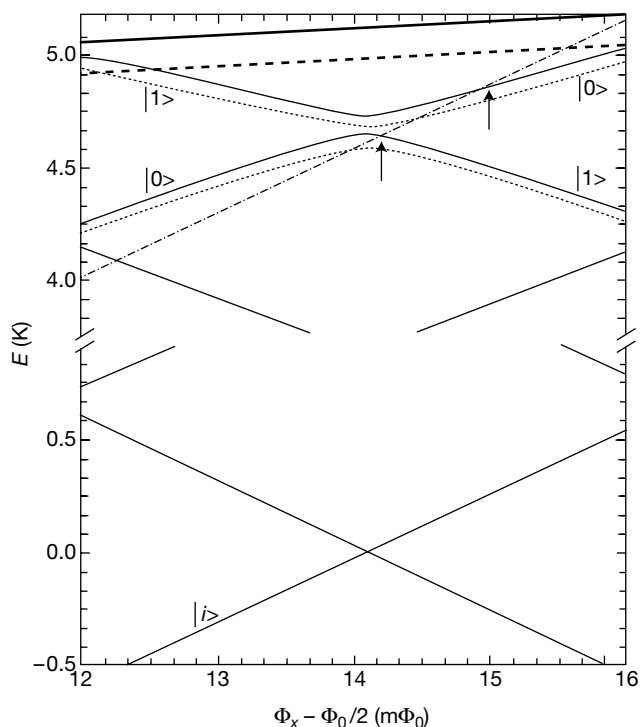


Figure 2 Calculated energy levels and photon-assisted tunnelling process. The thin solid lines represent the calculated SQUID energy levels as a function of Φ_x for a barrier height of $\Delta U_0 = 9.117$ K. The thick solid line is the calculated top of the energy barrier. The 96.0 GHz (4.61 K) applied microwaves boost the system out of the initial state $|i\rangle$, bringing it almost to the dot-dashed line. At certain values for Φ_x for which this line intersects one of the excited states (indicated by the arrows), a photon is absorbed and the system has a large probability of making an interwell transition. When the ΔU_0 is reduced (to 8.956 K), the levels and top of the barrier move down relative to $|i\rangle$ (dotted lines), changing the values of Φ_x at which photon absorption occurs. All energies are calculated relative to the mean energy of $|i\rangle$ and the lowest state of the right well; for clarity, the zero of energy is shifted to the point where the levels in the lower part of the figure cross.

A necessary condition for resolving this splitting is that the experimental linewidth of the states be smaller than Δ (ref. 16). The SQUID is extremely sensitive to external noise and dissipation (including that due to the measurement of Φ), both of which broaden the linewidth. Thus, the experimental challenges to observing coherent tunnelling are severe. The measurement apparatus must be weakly coupled to the system to preserve coherence, while the signal strength must be sufficiently large to resolve the closely spaced levels. In addition, the system must be well shielded from external noise. These challenges have frustrated previous attempts^{5,6} to observe coherence in SQUIDs.

The SQUID used in these experiments is made up of two Nb/AlO_x/Nb tunnel junctions in parallel, as shown in Fig. 1c; this essentially acts as a tunable junction in which I_c can be adjusted with an applied flux $\Phi_{x,d.c.}$. Thus, with Φ_x we control the tilt ϵ of the potential in Fig. 1a, while with $\Phi_{x,d.c.}$ we control ΔU_0 , the height of the energy barrier at $\epsilon = 0$. The flux state of our sample is measured by a separate d.c.-SQUID magnetometer inductively coupled to the sample. The sample is encased in a PdAu shield that screens it from unwanted radiation; a coaxial cable entering the shield allows the application of controlled external microwaves. The set-up is carefully filtered and shielded, as described elsewhere^{5,6}, and cooled to about 40 mK in a dilution refrigerator.

In our experiments, we probe the anticrossing of two excited levels in the potential by using microwaves to produce photon-assisted tunnelling. Figure 1a depicts this process for the case where the levels $|0\rangle$ and $|1\rangle$ are each localized in opposite wells. The system is initially prepared in the lowest state in the left well (labelled $|i\rangle$) with the barrier high enough that the rate for tunnelling out of $|i\rangle$ is negligible on the timescale of the measurement. Microwave radiation is then applied. When the energy difference between the initial state and an excited state matches the radiation frequency, the system has an appreciable probability of being excited into this state and subsequently decaying into the right well. This transition between wells results in a change in flux that can be detected by the magnetometer. Figure 2 shows the photon-assisted process when the excited levels have an anticrossing. From the SQUID's parameters (see below) and potential (equation (1)), it is straightforward to numerically solve for the energy levels of the SQUID in the zero-damping (completely coherent) limit. In Fig. 2 calculated levels (for $\Delta U_0 = 9.117$ K) are plotted as a function of Φ_x (thin solid lines). The calculated top of the barrier is indicated by the thick solid

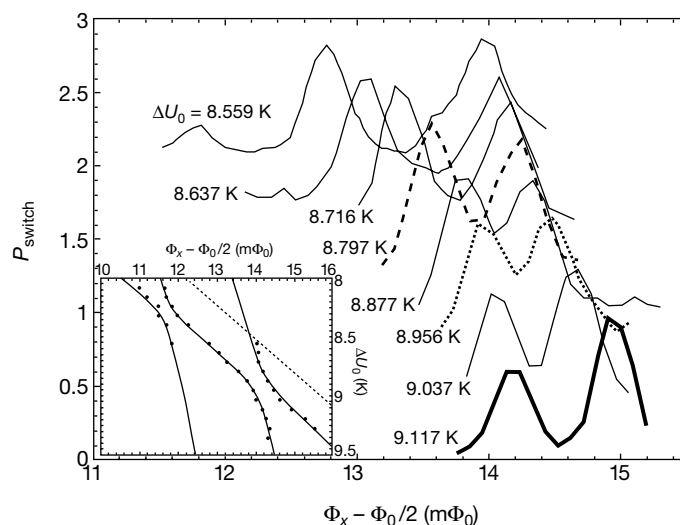


Figure 3 Experimental data. The main figure shows, as a function of Φ_x , the probability P_{switch} of making an interwell transition when a millisecond pulse of 96-GHz microwave radiation is applied. For clarity, each curve is shifted vertically by 0.3 relative to the previous one. As the energy barrier is reduced, the two observed peaks move closer

together and then separate: the signature of an anticrossing. The inset shows the position of the observed peaks in the ΔU_0 - Φ_x plane. Also shown is the calculated locus of points at which the virtual photon level (dot-dashed line in Fig. 2) intersects an excited level (solid lines) or the top of the classical energy barrier (dashed line).

line. The dot-dashed line represents level $|i\rangle$ shifted upward by the energy of the microwaves. At values of Φ_x for which this line intersects one of the excited levels (indicated by the arrows), the system can absorb a photon and make an interwell transition. When the barrier is reduced (to $\Delta U_0 = 8.956$ K), the excited levels and top of the barrier move to lower energy relative to $|i\rangle$ (dotted lines in the figure) and photon absorption occurs at different values of Φ_x . For a fixed frequency (a fixed frequency is used to ensure that the microwave power coupled to the sample remains constant for all resonances), we can map out the anticrossing by progressively reducing the barrier and thus moving the levels through the dashed photon line.

We use pulsed microwaves to excite the system to the upper levels. Before each pulse, the system is prepared in state $|i\rangle$ and the values of ΔU_0 and ϵ are set. Millisecond pulses of 96-GHz microwave radiation at a fixed power are applied and the probability of making a transition is measured. The experiment is repeated for various values of ϵ and ΔU_0 . Data from these measurements are shown in Fig. 3, where the probability of making a photon-assisted interwell transition is plotted as a function of Φ_x . Each curve, for a given ΔU_0 , is shifted vertically for clarity. Two sets of peaks are clearly seen. As ΔU_0 is decreased, these peaks move closer together and then separate without crossing. For $\Delta U_0 = 9.117$ K (thick solid curve), the right peak roughly corresponds to level $|0\rangle$, which is localized in the same well as $|i\rangle$ (compare with Fig. 2). The relative amplitude of the two peaks is due to the asymmetry of the potential (for this data, $|0\rangle$ is the fourth excited level in the left well and $|1\rangle$ is the tenth excited level in the right well) and is in accordance with recent calculations¹⁶. When ΔU_0 is decreased to 8.956 K (dotted curve), the peaks move closer together and the asymmetry disappears. The two states are now in the coherent regime and correspond approximately to the symmetric and anti-symmetric superpositions of the $|0\rangle$ and $|1\rangle$ states. As the barrier is decreased further (8.797 K is the dashed curve), the peaks move apart again and the asymmetry reappears, but now with a larger left peak and corresponding to $|0\rangle$. The two levels have thus passed through the anticrossing, changing roles without actually intersecting. The inset shows the positions of the peaks in the main figure (as well as other peaks) in the ΔU_0 - Φ_x plane; two anticrossings are clearly seen. The solid (dashed) lines in the inset represent the locus of points when the calculated energy levels (top of the barrier) are 96 GHz above the state $|i\rangle$. All of our data lie to the left of the dashed line and, therefore, correspond to levels that are below the top of the barrier. Hence, the flux-basis

states $|0\rangle$ and $|1\rangle$ are macroscopically distinct with mean fluxes that, we calculate, differ by about $\frac{1}{4}\Phi_0$. Each observed anticrossing thus represents the coherent superposition of macroscopically distinct states.

For one of the anticrossings, Fig. 4 shows the energy of the levels $|0\rangle$ and $|1\rangle$ as a function of ϵ relative to their mean energy, $E_{\text{mean}}(\Delta U_0, \Phi_x)$, calculated for each experimental point using the parameters listed below. This makes the data manifestly similar to Fig. 1b. At the middle of the anticrossing, the two levels have a tunnel splitting Δ of about 0.1 K in energy while the upper level is still about 0.15 K below the top of the classical energy barrier.

There are three parameters used for the calculations presented in Figs 3 and 4; $L, Z \equiv \sqrt{L/C}$ and β_L , all of which can be independently determined from measurements of classical phenomena or incoherent resonant tunnelling in the absence of radiation. From these independent measurements, we find $L = 240 \pm 15$ pH, $Z = 48.0 \pm 0.1 \Omega$ and $\beta_L = 2.33 \pm 0.01$. The values used in the calculation that yielded the best agreement with the data are $L = 238$ pH, $Z = 48 \Omega$ and $\beta_L = 2.35$, all in good agreement with the independently determined values.

We stress two related points regarding these results. First, at the anticrossing, both levels are below the top of the classical energy barrier. This is essential for the system to be in a superposition of macroscopically distinct flux states since the levels (in the absence of coherence) can only be associated with one well (one fluxoid state) if they are below the top of the barrier. The second point concerns the meaning of 'macroscopic'. The SQUID exhibits macroscopic quantum behaviour in two senses: (1) The quantum dynamics of the SQUID is determined by the flux through the loop, a collective coordinate representing the motion of approximately 10^9 Cooper pairs acting in tandem. Since the experimental temperature is about 500 times smaller than the superconducting energy gap, almost all microscopic degrees of freedom are frozen out and only the collective flux coordinate retains any dynamical relevance. (2) The two flux-basis states that we find to be superposed are macroscopically distinct. We calculate that for the anticrossings measured, the states $|0\rangle$ and $|1\rangle$ differ in flux by more than $\frac{1}{4}\Phi_0$ and differ in current by 2–3 μA . Given the geometry of the SQUID, this corresponds to a local magnetic moment of about $10^{10} \mu_B$, a truly macroscopic moment. $\square$

Received 12 April; accepted 8 June 2000.

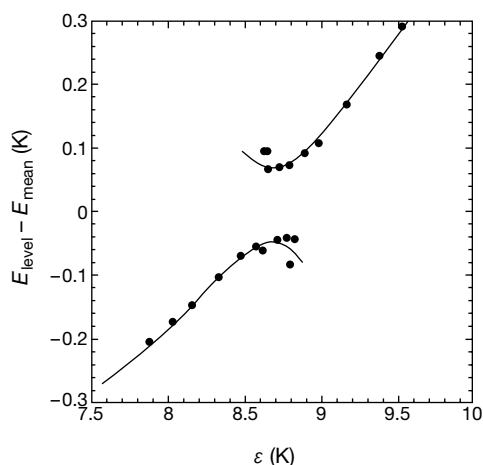


Figure 4 Energy of the measured peaks relative to the calculated mean of the two levels as a function of ϵ . At the midpoint of the figure, the measured tunnel splitting Δ between the two states in this anticrossing is about 0.1 K and the upper level is about 0.15 K below the top of the classical energy barrier. Calculated energy levels are indicated by the lines.

1. Schrödinger, E. Die gegenwärtige situation in der quantenmechanik. *Naturwissenschaften* **23**, 807–812, 823–828, 844–849 (1935).
2. Caldeira, A. O. & Leggett, A. J. Influence of dissipation on quantum tunneling in macroscopic systems. *Phys. Rev. Lett.* **46**, 211–214 (1981).
3. Leggett, A. J. *et al.* Dynamics of the dissipative 2-state system. *Rev. Mod. Phys.* **59**, 1–85 (1987).
4. Weiss, U., Grabert, H. & Linkwitz, S. Influence of friction and temperature on coherent quantum tunneling. *J. Low Temp. Phys.* **68**, 213–244 (1987).
5. Rouse, R., Han, S. & Lukens, J. E. Observation of resonant tunneling between macroscopically distinct quantum levels. *Phys. Rev. Lett.* **75**, 1614–1617 (1995).
6. Rouse, R., Han, S. & Lukens, J. E. in *Phenomenology of Unification from Present to Future* (eds Palazzi, G. D., Cosmelli, C. & Zanello, L.) 207–224 (World Scientific, Singapore, 1998).
7. Clarke, J., Cleland, A. N., Devoret, M. H., Esteve, D. & Martinis, J. M. Quantum mechanics of a macroscopic variable: the phase difference of a Josephson junction. *Science* **239**, 992–997 (1988).
8. Silvestrini, P., Palmieri, V. G., Ruggiero, B. & Russo, M. Observation of energy level quantization in underdamped Josephson junctions above the classical-quantum regime crossover temperature. *Phys. Rev. Lett.* **79**, 3046–3049 (1997).
9. Nakamura, Y., Pashkin, Y. A. & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
10. Friedman, J. R., Sarachik, M. P., Tejada, J. & Ziolo, R. Macroscopic measurement of resonant magnetization tunneling in high-spin molecules. *Phys. Rev. Lett.* **76**, 3830–3833 (1996).
11. del Barco, E. *et al.* Quantum coherence in Fe_8 molecular nanomagnets. *Europhys. Lett.* **47**, 722–728 (1999).
12. Wernsdorfer, W. *et al.* Macroscopic quantum tunneling of magnetization of single ferrimagnetic nanoparticles of barium ferrite. *Phys. Rev. Lett.* **79**, 4014–4017 (1997).
13. Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. A “Schrödinger cat” superposition state of an atom. *Science* **272**, 1131–1136 (1996).
14. Brune, M. *et al.* Observing the progressive decoherence of the “meter” in a quantum measurement. *Phys. Rev. Lett.* **77**, 4887–4890 (1996).
15. Arndt, M. *et al.* Wave-particle duality of C_{60} molecules. *Nature* **401**, 680–682 (1999).
16. Averin, D., Friedman, J. R. & Lukens, J. E. Macroscopic resonant tunneling of magnetic flux. Preprint cond-mat/0005081 at xxx.lanl.gov (2000).

Acknowledgements

We thank D. Averin and S. Han for useful conversations, J. Männik, R. Rouse and A. Lipski for technical advice and assistance and M. P. Sarachik for the loan of some equipment. This work was supported by the US Army Research Office and the US National Science Foundation.

Correspondence and requests for materials should be addressed to J.R.F. (e-mail: jonathan.friedman@sunysb.edu).

Generation and detection of phase-coherent current-driven magnons in magnetic multilayers

M. Tsoi*, A. G. M. Jansen*, J. Bass†, W.-C. Chiang†, V. Tsoi‡ & P. Wyder*

* Grenoble High Magnetic Field Laboratory, Max-Planck-Institut für Festkörperforschung and Centre National de la Recherche Scientifique, BP 166, F-38042 Grenoble Cedex 9, France

† Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824-1116, USA

‡ Institute of Solid State Physics RAS, 142432 Chernogolovka, Moscow Region, Russia

The magnetic state of a ferromagnet can affect the electrical transport properties of the material; for example, the relative orientation of the magnetic moments in magnetic multilayers¹ underlies the phenomenon of giant magnetoresistance. The inverse effect—in which a large electrical current density can perturb the magnetic state of a multilayer—has been predicted^{2–7} and observed experimentally with point contacts^{8,9} and lithographically patterned samples^{10,11}. Some of these observations were taken as indirect evidence for current-induced excitation of spin waves, or ‘magnons’. Here we probe directly the high-frequency behaviour and partial phase coherence of such current-induced excitations, by externally irradiating a point contact with microwaves. We determine the magnon spectrum and investigate how the magnon frequency and amplitude vary with the exciting current. Our observations support the feasibility of a spin-wave maser² or ‘SWASER’ (spin-wave amplification by stimulated emission of radiation).

A spin-momentum transfer associated with an electric current traversing a magnetic multilayer has been predicted^{2–7} to stimulate the emission of spin waves if the current density is large enough. In our experiments we inject current densities as high as 10^8 A cm⁻² into a multilayer through a point contact made by bringing a sharpened silver tip into contact with the multilayer film^{12,13}. Such a geometry, where a laterally unbounded multilayer is excited by a d.c. current whose distribution is localized by means of a point contact was shown⁷ to be equivalent to the case of a bounded multilayer^{2–3} of diameter comparable to that of the point contact. In both cases, spin-polarized currents induce a torque on magnetic moments in the ferromagnetic layers, thereby exciting their precession (spin waves). However, in the bounded multilayer this torque is applied to an entire magnetic layer, but in our case it is applied only to a small region of the layer just under the contact (the ‘disk’, for a circular contact), where the density of applied current is large enough. Spin-wave excitation is therefore expected only in this disk under the contact⁷. We⁸ and others⁹ have observed step increases in the static resistance ($R = V/I$) of such contacts for a certain bias current (see, for instance, equivalent peak structures in the derivative contact resistance dV/dI in Fig. 1b), attributable to the onset of magnon generation in the multilayer. Nevertheless, the main issue—the high-frequency nature of the excitation—was not conclusively addressed as magnetization reversal alone results in a similar increase in resistance^{9,11}.

To probe the frequencies of the current-driven magnons we have placed the point contact inside the microwave cavity of an electron spin resonance (ESR) spectrometer¹⁴, as shown schematically in Fig. 1a. Injection via a point contact of a high current density into the multilayer induces high-frequency precession of magnetization in the disk at frequency ω_1 . Using the ESR spectrometer to apply external radio-frequency radiation at frequency ω_2 , we expect to see a resonant amplification of this current-induced precession when ω_2 equals ω_1 . This process leads to an additional d.c. voltage across the contact that allows the frequency of precession to be mapped for different applied currents and fields. Here the contact serves as a lumped circuit element (much as rectifiers¹⁵ and frequency mixing diodes¹⁶ are used in radio engineering) that mixes the high-frequency response of the injected current (for example, radiation by current-driven magnons) with the external microwave radiation. The electromagnetic radiation couples into the contact junction through the wire tip, which acts as an antenna, thereby inducing high-frequency currents in the junction. For a point contact irradiated with an electromagnetic field that has components at two frequencies ω_1 and ω_2 , the junction current is

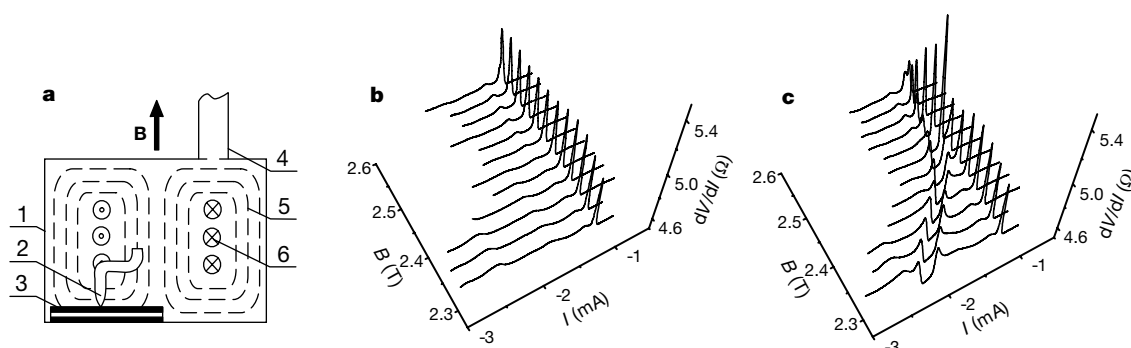


Figure 1 Experimental geometry and typical influence of the externally applied radiation on point-contact characteristics. **a**, Diagram of experimental geometry: 1, microwave cavity; 2, silver tip; 3, multilayer sample; 4, waveguide; 5 and 6 indicate magnetic and electric microwave field pattern. **b**, **c**, The point contact dV/dI spectra (solid lines) for a series of magnetic fields without (**b**) and with (**c**) external irradiation of the point contact with microwaves at $\omega/2\pi = 50.6$ GHz. Without irradiation (**b**) the spectra reveal just the usual peak structure for a certain negative bias current $I^*(B) \approx -1.2$ mA that we attribute

to the onset of the current-driven magnon excitations. Irradiation of the contact with microwaves (**c**) generates an additional structure in dV/dI . The strongest influence of the microwaves is found around a certain value $B^* \approx 2.6$ T of the applied field. Starting down from high fields, the additional structure is absent above B^* , appears around I^* when the field approaches B^* , and then moves to higher values of the bias current I as the field is reduced below B^* .

$I_{\text{tot}} = I + i_1 \cos(\omega_1 t + \varphi_1) + i_2 \cos(\omega_2 t + \varphi_2)$, where I is the d.c. current bias and $i_1 \cos(\omega_1 t + \varphi_1)$ and $i_2 \cos(\omega_2 t + \varphi_2)$ are the high-frequency currents induced by the radiation at frequencies ω_1 and ω_2 . The nonlinear mixing in the contact junction can be described by expanding the voltage $V(I_{\text{tot}})$ at the contact about the bias current I . For $i_1 = \text{const}$ (constant external radiation) and $i_1 \gg i_2(I)$ (bias-dependent radiation by current-driven magnons), the only significant radiation-induced d.c. voltage at the contact should develop when $\omega_1 = \omega_2$:

$$V_{\text{d.c.}} \propto i_1 (di_2/dI) (dV/dI) \cos(\varphi_1 - \varphi_2) \quad (1)$$

Our samples were sputtered $(\text{Co/Cu})_N$ multilayers with bilayer numbers N ranging from 20 to 50 and layer thicknesses $t_{\text{Co}} = 1.5$ nm and $t_{\text{Cu}} = 2.0$ – 2.2 nm, some having a 10-nm-thick iron underlayer (for details see ref. 8). No significant differences were found for different samples. A multilayer sample and a sharpened silver tip were placed in a microwave cavity at liquid helium temperature (4.2 K). After establishing an electrical contact between the tip and multilayer, its current–voltage (I – V) characteristics and their derivatives were recorded at different magnetic fields B applied perpendicular to the multilayer. Similar measurements were then repeated with the contact irradiated by microwaves within the frequency range of 40–60 GHz.

Figures 1b and c show the resulting point-contact spectra $dV/dI(I)$ (solid lines) for a series of magnetic fields without and with, respectively, the external irradiation of the point contact with microwaves at $\omega/2\pi = 50.6$ GHz. Without irradiation (Fig. 1b) the spectra reveal just the usual peak structure for a certain negative bias current $I^*(B) \approx -1.2$ mA. As noted above, such a peak structure corresponds to a step increase in $R = V/I$ that we attribute to the onset of the current-driven magnon excitations⁸ at I^* . In contrast, Fig. 1c shows that irradiation of the contact with microwaves generates an additional structure in dV/dI . The strongest influence of the microwaves is found around a certain value $B^* \approx 2.6$ T of the applied field. Starting down from high fields, the additional structure is absent above B^* , appears around I^* when the field approaches B^* , and then moves to higher values of the bias current I as the field is reduced below B^* . We also found the location of this additional radiation-induced structure to vary with the frequency ω of the applied radiation. Open symbols in Fig. 2 show the values of $I(B)$ where the radiation-induced structure appears for $\omega/2\pi = 43$ GHz (squares), 50.6 GHz (triangles), and 54.2 GHz (circles). The solid squares indicate the corresponding values of $B^*(\omega)$.

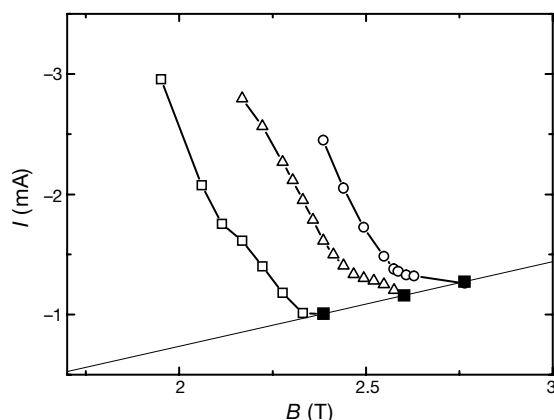


Figure 2 Phase diagram for the current-driven magnons. The solid line separates the two regions in the (I – B) current–magnetic field plane where the multilayer is unexcited (below the line) and excited (above the line). Open symbols show the values of $I(B)$ where the magnons of a given frequency $\omega/2\pi = 43$ GHz (squares), 50.6 GHz (triangles) and 54.2 GHz (circles) are excited in the multilayer (found from data as in Fig. 1). The solid squares indicate the corresponding values of $B^*(\omega)$ where the lowest frequency spin waves are excited for a given field.

We attribute this additional radiation-induced structure to rectification at the metal-to-metal point contact of the externally applied radiation and electromagnetic waves radiated by the current-driven magnons. As indicated by equation (1), this additional d.c. voltage is expected to appear when the frequency of the external microwaves equals the frequency of the magnons. Observation of the rectified signal implies that the current-driven magnons are at least partially coherent since an incoherent radiation would produce no signal at the contact due to the phase averaging indicated by the $\cos(\varphi_1 - \varphi_2)$ term in equation (1). Figure 2 represents a sort of phase diagram for the current-driven magnons, with the solid line separating the regions in the I – B plane where the multilayer is unexcited (below the line) and excited (above the line). The radiation-induced signal maps the values of I and B (open symbols in Fig. 2) at which the magnons of frequency equal to that of the external microwaves are excited in the multilayer. The negative slopes of the contours of constant ω (curves through the open symbols in Fig. 2) indicate that the wavenumber (k) of the current-induced magnons increases with increasing current, since for a fixed k at all currents these contours would be vertical straight lines.

For a given value of the applied magnetic field the lowest frequency spin waves are excited at the current threshold indicated by the solid line in Fig. 2. We have found their frequency to shift linearly with the applied magnetic field B^* , as shown in Fig. 3 (filled circles). For comparison, the open squares in Fig. 3 show the locations of the zero-wavenumber spin-wave resonances seen in the same sample. These latter points were found from measurements of the absorption of the applied microwaves in spin-wave resonance¹⁷ (see inset to Fig. 3) for which $\omega = \gamma(B_0 - \mu_0 M_s)$, where $\gamma = g\mu_B/\hbar$ is the gyromagnetic ratio, and M_s is the saturation magnetization. Figure 3 shows that B^* varies with ω in the same way as this resonance B_0 , but that $B^*(\omega)$ is smaller than $B_0(\omega)$ by about 0.6 T at the same ω . We attribute this difference to different polarizations (orientations of wavevector $\mathbf{k}$ relative to that of $\mathbf{M}_s$) of spin waves excited by the current and microwaves. We note that the microwave radiation excites the longitudinal spin waves ($\mathbf{k} \parallel \mathbf{M}_s$) and our multilayer set of very thin Co films behaves similarly (have the same g -factor and M_s) to bulk Co⁸. We suggest that the current excites the transverse spin waves ($\mathbf{k} \perp \mathbf{M}_s$) whose frequency $\omega_{\perp}$

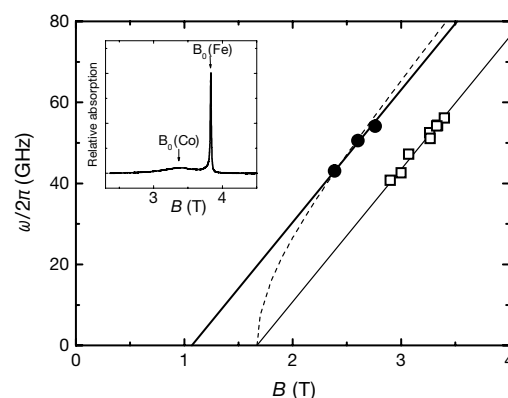


Figure 3 Frequency variation of the current-driven magnons with magnetic field. Filled circles show the frequency shift of the characteristic field B^* at which the magnons are excited by an electric current. For comparison, open squares show the frequency shift of the resonance field B_0 of the zero-wavenumber spin-wave mode (open squares) in the same sample (solid lines show the corresponding linear fits). The latter points were found in spin-wave resonance from the data, as in the inset, which shows a typical microwave absorption in the multilayer at $\omega/2\pi = 54.2$ GHz. A small peak at about 3.4 T is attributed to the excitation of the zero-wavenumber spin-wave mode in the thin Co layers of our multilayer. The strong resonance signal from Fe underlayer at about 3.9 T was used as a reference. The dashed line shows the frequency shift $\omega_{\perp}(B)$ of the transverse spin waves.

differs from that ($\omega_{\parallel}$) for $\mathbf{k} \parallel \mathbf{M}_S$. For small k values their ratio is given by $\omega_{\perp}/\omega_{\parallel} = (1 + \mu_0 M_S / (B - \mu_0 M_S))^{1/2}$ (ref. 18). Reconstruction of $\omega_{\perp}(B)$ using this equation and the frequency–field relation $\omega_{\parallel}(B)$ obtained in the spin-wave resonance (open squares in Fig. 3) gives the dashed curve in Fig. 3, which is consistent with our data for $B^*(\omega)$. The extremely small thickness of our Co layers (1.5 nm) suggests that the current excites transverse surface (not bulk) spin waves whose frequency variation versus field is similar to that of $\omega_{\perp}(B)$ (ref. 19).

Equation (1) allows the amplitude $i_2(I)$ of the current-driven magnons to be reconstructed from the shape of the radiation-induced structure. An example of such a reconstruction is shown in Fig. 4. Figures 4a and b show the derivative contact resistance dV/dI and the static contact resistance $R = V/I$, respectively, as functions of the bias current I . The squares (circles) and black (grey) trace show the point-contact spectra with the external microwave radiation off (on). Since the additional radiation-induced structure at $I_c \approx 2$ mA is well shifted from I^* , where we see strong variations in dV/dI , the latter can be approximated in equation (1) by a constant. Figure 4c shows $i_2(I)$ found by integration of the additional radiation-induced d.c. voltage around I_c . This curve represents the relative amplitude of radiation by current-driven magnons at 50.6 GHz. Starting from low bias currents, this amplitude is zero, rises to a maximum at I_c , and then decreases to a constant value beyond I_c . The maximum at I_c can be attributed to amplification of the magnon generation induced by an external radiation of the same frequency. This observation supports the feasibility of spin-wave amplification by

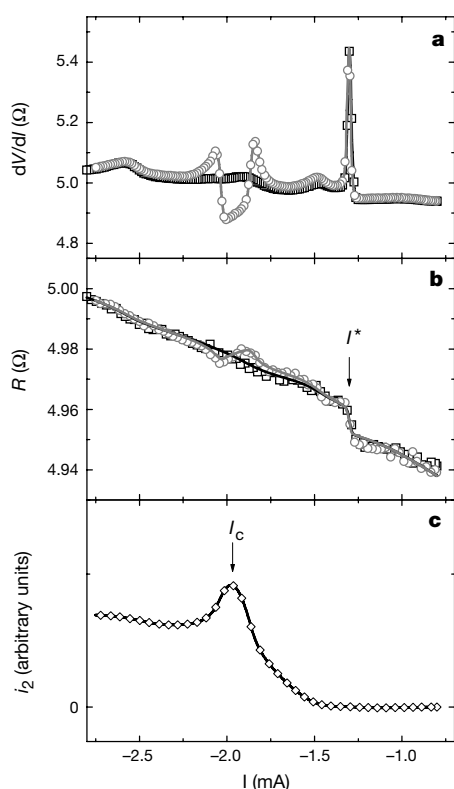


Figure 4 Variation of the amplitude of current-driven magnons with the exciting current. **a, b**, The derivative contact resistance dV/dI and the static contact resistance $R = V/I$, respectively, as functions of the bias current I (magnetic field $B = 2.3$ T). The squares (circles) and black (grey) trace show the point-contact spectra with the external microwave radiation off (on). The peak structure at the threshold current I^* corresponds to the onset of the current-driven excitations. The external radiation at 50.6 GHz induces an additional structure around the critical current I_c . **c**, The relative amplitude of radiation by current-driven magnons $i_2(I)$ at 50.6 GHz found by integration of the additional radiation-induced d.c. voltage around I_c .

stimulated emission of radiation (SWASER²), whereas the observed magnon polarization is of importance for the configuration of a possible SWASER. □

Received 10 January; accepted 5 May 2000.

1. Baibich, M. N. *et al.* Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. *Phys. Rev. Lett.* **61**, 2472–2475 (1988).
2. Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353–9358 (1996).
3. Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1–L7 (1996).
4. Berger, L. Multilayer as spin-wave emitting diodes. *J. Appl. Phys.* **81**, 4880–4882 (1997).
5. Bazaliy, Ya. B., Jones, B. A. & Zhang, S.-C. Modification of the Landau-Lifshitz equation in the presence of a spin-polarized current in colossal- and giant-magnetoresistive materials. *Phys. Rev. B* **57**, R3213–R3216 (1998).
6. Berger, L. Spin-wave emitting diodes and spin diffusion in magnetic multilayers. *IEEE Trans. Magn.* **34**, 3837–3841 (1998).
7. Slonczewski, J. C. Excitation of spin waves by an electric current. *J. Magn. Magn. Mater.* **195**, L261–L268 (1999).
8. Tsoi, M. *et al.* Excitation of a magnetic multilayer by an electric current. *Phys. Rev. Lett.* **80**, 4281–4284 (1998).
9. Myers, E. B., Ralph, D. C., Katine, J. A., Louie, R. N. & Buhrman, R. A. Current-induced switching of domains in magnetic multilayer devices. *Science* **285**, 867–870 (1999).
10. Katine, J. A., Albert, F. J., Buhrman, R. A., Myers, E. B. & Ralph, D. C. Current-driven magnetization reversal and spin wave excitations in Co/Cu/Co pillars. Preprint cond-mat/9908231 at (xxx.lanl.gov) (1999).
11. Sun, J. Z. Current-driven magnetic switching in manganite trilayer junctions. *J. Magn. Magn. Mater.* **202**, 157–162 (1999).
12. Jansen, A. G. M., van Gelder, A. P. & Wyder, P. Point-contact spectroscopy in metals. *J. Phys. C* **13**, 6073–6118 (1980).
13. Tsoi, M., Jansen, A. G. M. & Bass, J. Search for point-contact giant magnetoresistance in Co/Cu multilayers. *J. Appl. Phys.* **81**, 5530–5532 (1997).
14. Seck, M. & Wyder, P. A sensitive broadband high-frequency electron spin resonance/electron nuclear double resonance spectrometer operating at 5–7.5 mm wavelength. *Rev. Sci. Instrum.* **69**, 1817–1822 (1998).
15. van der Heijden, R. W., Jansen, A. G. M., Stoelinga, J. H. M., Swartjes, H. M. & Wyder, P. A new mechanism for high-frequency rectification at low temperatures in point contacts between identical metals. *Appl. Phys. Lett.* **37**, 245–248 (1980).
16. Javan, A. in *Fundamental and Applied Laser Physics* (eds Field, M. S., Javan, A. & Kurnit, N. A.) 295–334 (Wiley, New York, 1973).
17. Seavey, M. H. Jr & Tannenwald, P. E. Direct observation of spin-wave resonance. *Phys. Rev. Lett.* **1**, 168–169 (1958).
18. Herring, C. & Kittel, C. On the theory of spin waves in ferromagnetic media. *Phys. Rev.* **81**, 869–880 (1951).
19. Damon, R. W. & Eshbach, J. R. Magnetostatic modes of a ferromagnet slab. *J. Phys. Chem. Solids* **19**, 308–320 (1961).

Correspondence and requests for materials should be addressed to M. T. (e-mail: tsoim@labs.polycnrs-gre.fr).

Acknowledgements

This work was supported in part by the NSF, RU MNT, and RFFI.

Self-directed growth of molecular nanostructures on silicon

G. P. Lopinski, D. D. M. Wayner & R. A. Wolkow

Steacie Institute for Molecular Sciences, National Research Council, 100 Sussex Drive, Ottawa, Ontario, Canada, K1A 0R6

Advances in techniques for the nanoscale manipulation of matter are important for the realization of molecule-based miniature devices^{1–8} with new or advanced functions. A particularly promising approach involves the construction of hybrid organic-molecule/silicon devices^{9–14}. But challenges remain—both in the formation of nanostructures that will constitute the active parts of future devices, and in the construction of commensurately small connecting wires. Atom-by-atom crafting of structures with scanning tunnelling microscopes^{15–17}, although essential to fundamental advances, is too slow for any practical fabrication

process; self-assembly approaches may permit rapid fabrication¹⁸, but lack the ability to control growth location and shape. Furthermore, molecular diffusion on silicon is greatly inhibited¹⁹, thereby presenting a problem for self-assembly techniques. Here we report an approach for fabricating nanoscale organic structures on silicon surfaces, employing minimal intervention by the tip of a scanning tunnelling microscope and a spontaneous self-directed chemical growth process. We demonstrate growth of straight molecular styrene lines—each composed of many organic molecules—and the crystalline silicon substrate determines both the orientation of the lines and the molecular spacing within these lines. This process should, in principle, allow parallel fabrication of identical complex functional structures.

On a clean Si(100) surface, an alkene reacts with the dangling bonds of a Si dimer to form two Si–C bonds, in a process analogous to a [2+2] cycloaddition^{20–24}. We force a different process to ensue by allowing an alkene to react with only one Si dangling bond on an otherwise hydrogen-terminated surface. Initially, the C–C π -bond opens to form one Si–C bond and one C radical, as shown schematically in Fig. 1. If subsequently a hydrogen atom is abstracted from a neighbouring surface site to satisfy the radical, a new Si dangling bond—and the possibility for a chain reaction—is created. Such a mechanism has been suggested to account for alkene reactions with H-terminated Si(111) under liquid-phase reaction conditions¹³. On H-terminated Si(111), which presents a hexagonal array of sites, the chain reaction is expected to lead to irregular islands of adsorbed molecules²⁵. On the anisotropic Si(100) surface, a directional preference associated with the H-abstraction step may be expected, directing the growth of one-dimensional lines.

Measurements were performed in ultrahigh vacuum. Clean Si(100) surfaces were obtained by heating As-doped Si crystals (0.005 Ω cm) to 1,250 °C. The surfaces were then H-terminated by exposure to atomic hydrogen at a temperature of 330 °C (ref. 26).

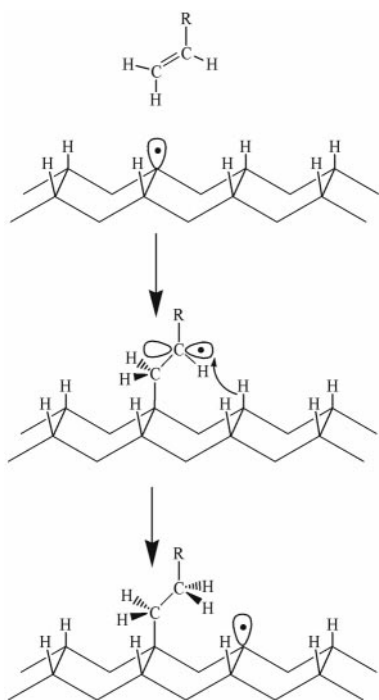


Figure 1 Proposed chain reaction mechanism for self-directed growth of molecular nanostructures on silicon. This involves the reaction of an alkene with a single dangling bond on an otherwise hydrogen-terminated Si(100) surface. The initial reaction involves formation of a carbon-centred radical that can then abstract a hydrogen from an adjacent dimer along a row, creating a new Si dangling bond. Changing the R group from methyl (propylene) to phenyl (styrene) stabilizes the intermediate state, increasing the probability of the chain reaction.

Single Si dangling bonds were created with the tip of the scanning tunnelling microscope¹⁶ (STM).

Our initial attempt at growing molecular lines via the scheme outlined in Fig. 1 employed propylene. After exposure of over 200 L (1 Langmuir (L) = 10^{-6} torr s), isolated adsorption events could be seen but the intended chain reaction had not occurred. This can be rationalized by noting the weak adsorption energy of the radical intermediate. Density functional theory calculations (B3LYP/6-31G* employing a 15 Si atom cluster) indicate that this state is bound by only 0.4 eV, implying a residence time of ~ 1 μ s at room temperature. Hydrogen abstraction is a favourable process, calculated to lower the total energy by a further 0.7 eV, but if the residence time of the radical species is too short, desorption will occur before abstraction. An adjacent phenyl ring is known to stabilize a C radical, suggesting styrene as a good candidate for observing a chain reaction. The expectation of increased radical stability is confirmed by a calculated binding energy of 0.8 eV, implying a residence time of ~ 10 s. Hydrogen abstraction by the styrene radical is accompanied by a further 0.1-eV energy release.

Figure 2 shows a sequence of STM images of an H-terminated Si(100) surface with a number of single dangling Si bonds upon increasing doses of styrene. After a 3-L exposure, some unreacted dangling bonds remain, and are imaged as ~ 1 -Å-high protrusions. At other sites, adsorption has occurred, as indicated by the emergence of distinctly larger protrusions ranging in height from 2 to 3 Å. Further exposure leads to several more of the dangling-bond sites reacting, as well as to development of some short lines extending over 2 or 3 surface sites (Fig. 2b). Still further exposure caused continued growth of some of the lines. The lines tend to appear centred between dimer rows. These observations are consistent with the radical chain mechanism sketched above. After initial reaction at a dangling-bond site, H abstraction occurs preferentially from an adjacent dimer, leading to growth of a line along one side of a dimer row. The longest line in Fig. 2d is ~ 130 Å long, corresponding to 34 adsorption sites. Growth of the lines is often observed to be stopped at pre-existing 'missing dimer' (or other) defects, as seen for the line indicated in the top left-hand corner of Fig. 2d. We note the tip has no role in the growth process—line growth could be observed by dosing while scanning or with the tip retracted.

The growth kinetics of these styrene lines exhibits several interesting aspects. Figure 3a shows an STM image of a number of styrene lines grown on a H-terminated surface with a dilute

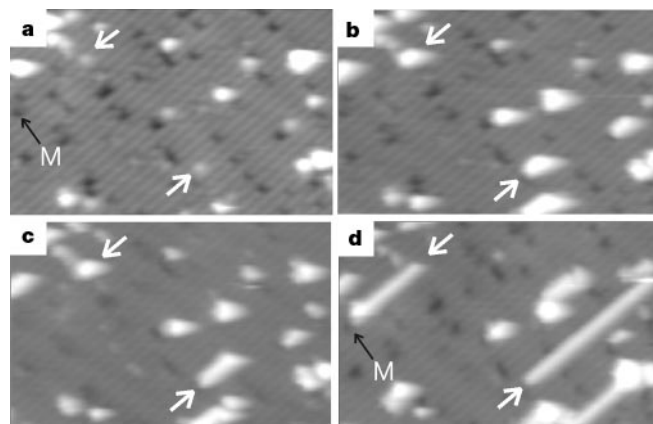


Figure 2 Growth of styrene lines on a H-terminated Si(100) surface with a dilute concentration of single Si dangling bonds. The figure shows a sequence of STM images ($250 \text{ Å} \times 140 \text{ Å}$, -2.1 V, 47 pA) corresponding to an increasing exposure to styrene: **a**, 3 L; **b**, 28 L; **c**, 50 L; and **d**, 105 L. The white arrows denote two particular dangling-bond sites that lead to the growth of long styrene lines. The missing dimer defect (M) marked in the figure terminates the growth of the line in the top left-hand corner of the image.

concentration of single Si dangling bonds. Some lines grow much longer than others; indeed, some dangling bonds exhibit no growth at all. This is not what is expected if the probability of an additional growth event is independent of chain length. In fact, longer chains are more numerous than expected on this basis, indicating that there is some preferential growth of longer chains. To account for this autocatalytic effect, we note that styrene adsorption is probably mediated by a precursor state that allows rather free diffusion over the H-terminated surface. We suggest that the physically adsorbed precursor molecules preferentially adhere to existing molecular lines, thereby creating increased concentration and growth in that vicinity. Monte Carlo simulations may help unravel the details of growth²⁷. From Fig. 3a it is also evident that, in addition to single rows of styrene molecules, some double lines are also observed. These result from rare abstraction events perpendicular to the dimer row direction, involving the other H atom within a dimer or on the adjacent row. This can result in either a jog in the molecular line with growth continuing in the same direction or the line doubling back on itself, growing back down the other side of the dimer or along the adjacent row. Although both of these are observed, double lines are observed to be more likely, consistent with the terminal radical adsorbate leaning toward neighbouring adsorbates, thereby giving a bias to the H-abstraction event.

In Fig. 3b an image revealing some molecular structural detail is shown. The individual protrusions are spaced by 3.8 Å, corresponding to the distance between dimers on the Si(100) surface. This image is consistent with molecules bound to adjacent dimers on the same side of a row, as shown in Fig. 3b inset. While the conformation of the adsorbed molecules cannot be determined from the STM images, the alignment of the phenyl rings is important for determining the electronic properties of this line of molecules. If the rings are parallel as in the inset, some degree of overlap between π -states is expected, suggesting that the lines of molecules could function as molecular wires. Arrangements of this kind are known to be effective hole conductors. The intermolecular spacing of 3.8 Å is comparable to that in molecular crystals such as metal phthalocyanines and polycyclic aromatic hydrocarbons, which show substantial effects of intermolecular π -coupling. Furthermore, extended Hückel calculations of the lowest unoccupied molecular orbitals for this structure indicate considerable electronic coupling between adjacent molecules. Though the present lines may or may not have the qualities of wires, the H-abstraction method for growing nanostructures is not limited to styrene. We anticipate that various alkenes, alkynes and other molecules may be used to design-in molecular wire (or insulator, or specific chemical adsorption site, or other) properties.

An alternative explanation for the linear structures described here—that each line is a single oligostyrene molecule attached to the surface by only one bond—must also be considered. Like the

H-abstraction mechanism, styrene polymerization is a radical chain reaction. It is conceivable that, upon formation of a styrene radical, polymerization with a second styrene molecule occurs, rather than surface H abstraction. Only the H-abstraction scheme, however, accounts for all of the experimental observations. It is unlikely that the weak van der Waals interaction between the molecule and the surface, of the order of 0.1 eV per molecule (monomer), would hold a polymer straight and in registry with the dimer rows of the surface. We note, furthermore, that no displacements of the lines have been observed even when the tip was brought very close to the molecules—conditions we expect would sweep molecules aside if only physically adsorbed. The fact that the intermolecular spacing along a line is determined by the distance between Si dimers is also difficult to reconcile with polystyrene formation in which the distance between phenyl rings is 2.52 Å. Another point in favour of H abstraction is the observation that line growth has not been observed to continue over defects. While polystyrene growth may be expected to proceed unimpeded, it is reasonable that a defect lacking an H atom for abstraction, or that does not allow formation of a Si dangling bond, could stop growth via the H-abstraction mechanism. Furthermore, one can reasonably account for the formation of double lines only by the H-abstraction scheme. Last, it is notable that under our experimental conditions the maximum collision rate of styrene with the carbon radical is $\sim 1 \text{ s}^{-1}$, a rate many orders lower than in the liquid phase where styrene polymerization is normally observed. Based on the known rate constant for styrene polymerization in solution²⁸, the rate of polymer formation in our experiments is expected to be less than $\sim 4 \times 10^{-9} \text{ s}^{-1}$, much too low to account for the observed line growth.

It is important to note that molecular lines grown via the H-abstraction mechanism are expected to be unstable. A line ending in a Si dangling bond can ‘unzip’ if the terminal molecule reverses the abstraction reaction, donating its hydrogen back to the dangling bond at the end of the line followed by desorption of styrene. The fact that chain growth is observed indicates that the reverse reaction is quite slow. Rare unzipping events have been observed at room temperature after ~ 20 hours, yielding an estimated barrier for the reverse reaction of 1.15 eV. Once the dangling bond at the end of the line is removed, the stability of the line should be increased significantly. This might be achieved in a number of ways; for example, by exposure to atomic hydrogen. Without intentional termination measures, a minority of lines are observed to survive heating up to 300 °C, which is unexpected given the estimated 1.15-eV barrier for the reverse reaction. We suggest that these lines have been stabilized by an opportunistic termination event involving a surface defect that quenches either the carbon or silicon radical. For example, if abstraction occurs on a dimer with a pre-existing dangling bond, that pair of dimers will be quickly consumed by di- σ addition, terminating the chain reaction.

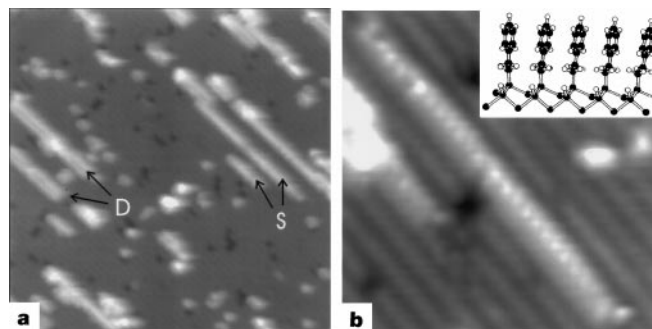


Figure 3 Structure of styrene lines. **a**, An STM image (350 Å^2 , -3.3 V , 10 pA) of a H-terminated Si(100) surface with a dilute concentration of single Si dangling bonds after exposure to 200 L of styrene. Examples of single (S) and double lines (D) are indicated. **b**, A molecularly resolved STM image (90 Å^2 , -2.2 V , 470 pA) of a single styrene line. The

features in the line are spaced by $\sim 3.8 \text{ Å}$. The inset depicts a possible conformation of a five-molecule chain of adsorbed styrene molecules, each of which is covalently bound to a single Si dimer. The parallel alignment of the phenyl rings in this structure is expected to lead to overlap between the electronic states of the individual molecules.

On the basis of preliminary observations, we consider that—although many parameters remain to be explored—the method outlined here for self-directed growth of molecular nanostructures on silicon could be sufficiently controllable to be useful in forming and/or connecting nanostructures. In particular, we note that by combining the chain reaction mechanism demonstrated here with state-of-the-art tip-induced H-removal methods²⁹ and preparation methods for Si(100) surfaces with extremely low defect densities³⁰ recently demonstrated by others, it should be possible to join pre-defined points with an extended molecular line. By appropriate design of reactant molecules and seed conditions it is, in principle, possible to control growth direction along a row, to turn corners and to grow vertically, to create complex, functional structures. Furthermore, because many seed sites may be exposed to reactant molecules simultaneously, this approach may allow the parallel fabrication of arrays of identical structures by simply opening and closing valves. □

Received 14 December 1999; accepted 15 May 2000.

- Aviram, A. & Ratner, M. A. Molecular rectifiers. *Chem. Phys. Lett.* **29**, 277–283 (1974).
- Tian, W. *et al.* Conductance spectra of molecular wires. *J. Chem. Phys.* **109**, 2874–2882 (1998).
- Reed, M. A., Zhou, C., Muller, C. J., Burgin, T. P. & Tour, J. M. Conductance of a molecular junction. *Science* **278**, 252–254 (1997).
- Collier, C. P. *et al.* Electronically configurable molecular-based logic gates. *Science* **285**, 391–394 (1999).
- Tans, S. J., Verschueren, A. R. M. & Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
- Davis, W. B., Svec, W. A., Ratner, M. A. & Wasielewski, M. R. Molecular wire behavior in *p*-phenylenevinylene oligomers. *Nature* **396**, 60–63 (1998).
- Ness, H. & Fisher, A. J. Quantum inelastic conductance through molecular wires. *Phys. Rev. Lett.* **83**, 452–455 (1999).
- Joachim, C. & Gimzewski, J. K. An electromechanical amplifier using a single molecule. *Chem Phys. Lett.* **265**, 353–355 (1997).
- Wolkow, R. A. Controlled molecular adsorption on Si: laying a foundation for molecular devices. *Annu. Rev. Phys. Chem.* **50**, 413–441 (1999).
- Hamers, R. J., Hovis, J. S., Lee, S., Liu, H. & Shan, J. Formation of ordered, anisotropic organic monolayers on the Si surface. *J. Phys. Chem.* **101**, 1489–1492 (1997).
- Wolkow, R. A., Lopinski, G. P. & Moffatt, D. J. Resolving organic molecule-silicon scanning tunneling microscopy features with molecular orbital methods. *Surf. Sci.* **416**, L1107–L1113 (1998).
- Lopinski, G. P., Moffatt, D. J., Wayner, D. D. M. & Wolkow, R. A. Determination of the absolute chirality of individual adsorbed molecules using the scanning tunnelling microscope. *Nature* **392**, 909–911 (1998).
- Linford, M. R., Fenter, P., Eisenberger, P. M. & Chidsey, C. E. D. Alkyl monolayers on silicon prepared from 1-alkenes and hydrogen-terminated silicon. *J. Am. Chem. Soc.* **117**, 3145–3155 (1995).
- Cohen, R., Zenou, N., Cahen, D. & Yitzchaik, S. Molecular electronic tuning of Si surfaces. *Chem Phys. Lett.* **279**, 270–274 (1997).
- Eigler, D. M. & Schweizer, E. K. Positioning single atoms with a scanning tunnelling microscope. *Nature* **344**, 524–526 (1990).
- Avouris, Ph. *et al.* Atom-scale desorption through electronic and vibrational excitation mechanisms. *Surf. Sci.* **363**, 368–377 (1996).
- Shen, T. C., Wang, C. & Tucker, J. R. Al nucleation on monohydride and bare Si(001) surfaces: atomic scale patterning. *Phys. Rev. Lett.* **78**, 1271–1274 (1997).
- Nogami, J. in *Atomic and Molecular Wires* (eds Joachim, C. & Roth, S.) 11–21 (Kluwer, Dordrecht, 1997).
- Wolkow, R. A. & Moffatt, D. J. The frustrated motion of benzene on the surface of Si(111). *J. Chem. Phys.* **103**, 10696–10700 (1995).
- Bozack, M. J., Taylor, P. A., Choyke, W. J. & Yates, J. T. Chemical activity of the C = C double bond. *Surf. Sci.* **177**, 933–937 (1986).
- Yoshinobu, J., Tsuda, H., Onchi, M. & Nishijima, M. The adsorbed states of ethylene on Si(100)(c(4x2), Si(100)(2x1), and vicinal Si(100)9°. Electron energy loss spectroscopy and low energy electron diffraction studies. *J. Chem. Phys.* **87**, 7332–7340 (1987).
- Mayne, A. J. *et al.* An STM study of the chemisorption of C₂H₄ on Si(001)(2x1). *Surf. Sci.* **284**, 247–256 (1993).
- Pan, W., Zhu, T. & Yang, W. First-principles study of the structure and electronic properties of ethylene adsorption on Si(100)-2x1 surface. *J. Chem. Phys.* **107**, 3981–3985 (1997).
- Fisher, A. J., Blochl, P. E. & Briggs, G. A. D. Hydrocarbon adsorption on Si(001): When does the silicon dimer bond break? *Surf. Sci.* **374**, 298–305 (1997).
- Cicero, R. L., Linford, M. R. & Chidsey, C. E. D. Photoreactivity of unsaturated compounds with hydrogen terminated silicon (111). *Langmuir* (in the press).
- Boland, J. J. Role of bond strain in chemistry of hydrogen on the Si(100) surface. *Surf. Sci.* **261**, 17–28 (1992).
- Mo, W., Kleiner, J., Webb, M. B. & Lagally, M. G. Activation energy for surface diffusion of Si on Si(100): A scanning tunneling microscopy study. *Phys. Rev. Lett.* **66**, 1998–2001 (1991).
- O'Driscoll, K. F. & Mahabadi, H. K. Spatially intermittent polymerization. *J. Polymer Sci. Chem.* **14**, 869–881 (1976).
- Hitosugi, T. *et al.* Jahn-Teller distortion in dangling-bond linear chains fabricated on a hydrogen-terminated Si(100)-2 × 1 surface. *Phys. Rev. Lett.* **82**, 4034–4037 (1999).
- Hata, K., Kimura, T., Ozawa, S. & Shigekawa, K. How to fabricate a defect free Si(100) surface. *J. Vac. Sci. Technol.* (in the press).

Correspondence and requests for materials should be addressed to R.A.W. (e-mail: bob.wolkow@nrc.ca).

Direct measurement of hole transport dynamics in DNA

Frederick D. Lewis, Xiaoyang Liu, Jianqin Liu, Scott E. Miller, Ryan T. Hayes & Michael R. Wasielewski

Department of Chemistry, Northwestern University, Evanston, Illinois 60208-3113, USA

Our understanding of oxidative damage to double helical DNA and the design of DNA-based devices for molecular electronics is crucially dependent upon elucidation of the mechanism and dynamics of electron and hole transport in DNA^{1–4}. Electrons and holes can migrate from the locus of formation to trap sites^{1,5}, and such migration can occur through either a single-step “super-exchange” mechanism or a multistep charge transport “hopping” mechanism^{6–17}. The rates of single-step charge separation and charge recombination processes are found to decrease rapidly with increasing transfer distances^{6–9}, whereas multistep hole transport processes are only weakly distance dependent^{10–13}. However, the dynamics of hole transport has not yet been directly determined. Here we report spectroscopic measurements of photoinduced electron transfer in synthetic DNA that yield rate constants of $\sim 5 \times 10^7 \text{ s}^{-1}$ and $5 \times 10^6 \text{ s}^{-1}$, respectively, for the forward and return hole transport from a single guanine base to a double guanine base step across a single adenine. These rates are faster than processes leading to strand cleavage, such as the reaction of guanine cation radical with water², thus permitting holes to migrate over long distances in DNA. However, they are too slow to compete with charge recombination in contact ion pairs^{6,7}, a process which protects DNA from photochemical damage.

Our approach to the investigation of distance-dependent electron transfer in DNA utilizes hairpin-forming bis(oligonucleotide) conjugates in which a stilbene-4,4'-dicarboxamide chromophore

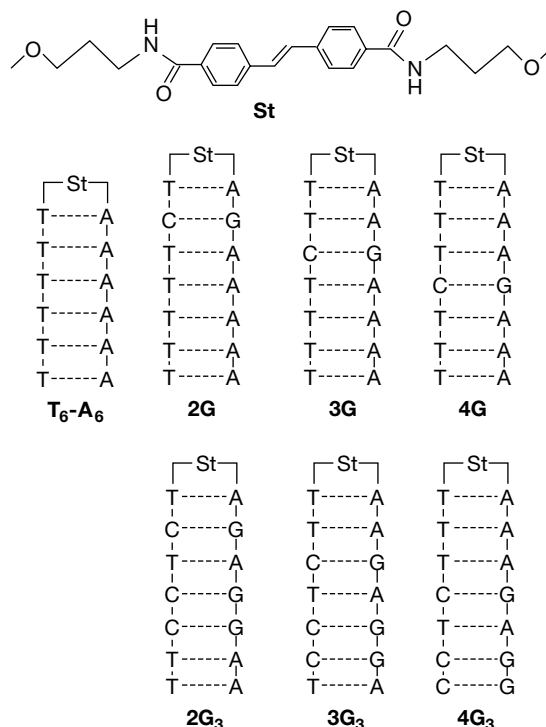


Figure 1 Structures of the stilbene linker and synthetic hairpins. T --- A and C --- G are thymine:adenine and cytosine:guanine base pairs, respectively.

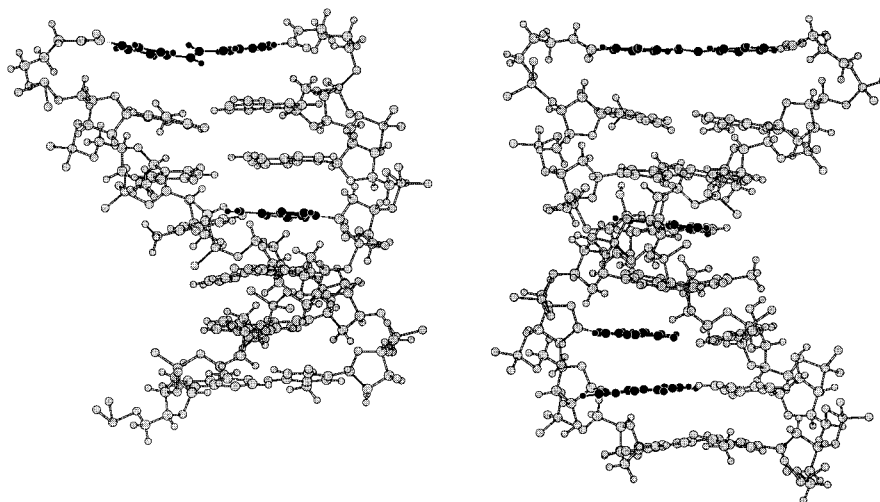


Figure 2 Structures for the hairpins 3G (panel **a**) and 3G₃ (panel **b**). The stilbenedicarboxamide linker (top of structure) and all guanines are darkened. Structures were calculated using the molecular mechanics force field (MM⁺) within Hyperchem

(St) serves as a linker between two complementary oligonucleotides (Fig. 1)^{6,7}. The singlet stilbene selectively photooxidizes guanine, but not the other three nucleobases, resulting in the formation of the stilbene anion radical (St^{-•}) and guanine cation radical (G^{+•}). By monitoring the formation and decay of St^{-•}, the dynamics of charge separation and charge recombination can be determined by transient spectroscopy over a large dynamic range (sub-picosecond to microsecond). Comparison of the results for hairpins containing a single G:C base pair with those for the parent hairpin (2G, 3G or 4G versus T₆-A₆, Fig. 1) has shown that the dynamics of electron transfer processes are strongly dependent upon the number of A:T base pairs separating the stilbene acceptor and guanine donor. These results are in accord with a single-step, superexchange mechanism for electron transfer, according to which $k_{et} = k_0 \exp(-\beta R)$, where k_0 is a pre-exponential factor, R is the distance between the donor and acceptor, and the parameter β is a characteristic of the bridge and its coupling with D and A (ref. 15). The values of β derived from direct kinetic measurements of the formation (charge separation) and decay (charge recombination) of St^{-•} are $0.7 \pm 0.1 \text{ \AA}^{-1}$ and $0.9 \pm 0.1 \text{ \AA}^{-1}$, respectively^{6,7}. When a single G is replaced by a GG step, the rate constant for charge separation is about twice as fast and that for charge recombination is about three times slower, indicating that the GG step serves as a shallow hole trap⁷.

The structure of the stilbene-linked synthetic DNA hairpins can be readily modified to include additional G:C base pairs. Using standard solid-state synthetic methodology¹⁸, hairpins containing three G:C base pairs (Fig. 1) have been prepared and characterized by absorption, fluorescence, and mass spectroscopy. The identifying

V5.01a (Hypercube, Waterloo, Ontario, 1997). Local minima were optimized assuming normal B-form DNA geometries.

symbols in Fig. 1 specify the location of the first guanine with respect to the stilbene linker (numerical prefix) and the total number of guanines in the hairpin (subscript). Molecular modelling of the stilbene-linked hairpins indicates that they adopt B-form structures in which the stilbene is approximately parallel to the adjacent base pair, as shown in Fig. 2 for the hairpins 3G and 3G₃. The crystal structure of a stilbenediether-linked hairpin has recently been reported, and is similar to that provided by molecular modelling of either the stilbenediether or the diamide linker¹⁹.

The hairpins that contain three G:C base pairs were designed to probe the dynamics of hole transport from G^{+•}, formed upon photoinduced charge separation, to a GG step separated from G^{+•} by one T:A base pair. The transient absorption spectra and kinetics of the stilbene-linked hairpins at times < 6 ns were measured with a femtosecond amplified Ti:sapphire laser system (340-nm excitation, 200-fs total instrument response function), while such measure-

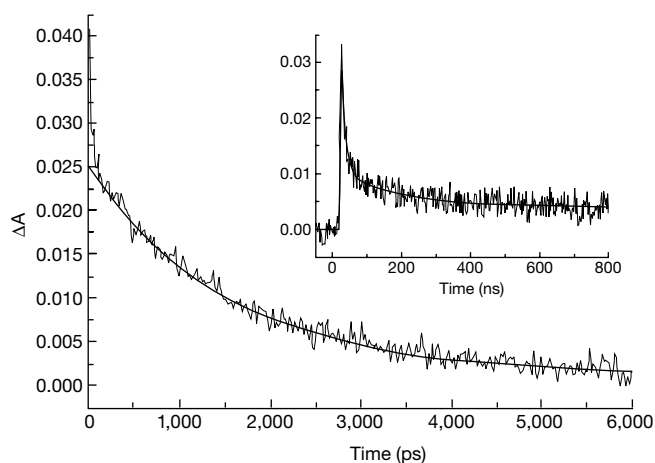


Figure 3 Decay of the transient absorption spectrum of 3G₃ probed at 575 nm. Data are shown for picosecond (main figure) and nanosecond (inset) timescales. The smooth curve superimposed on the ps decay curve is the fit to the kinetic model depicted in Fig. 4 using $k_1 = 5.0 \times 10^7 \text{ s}^{-1}$ and $k_{-1} = 6.0 \times 10^6 \text{ s}^{-1}$. The ~15-ps component present in the ps data is due to relaxation within the stilbene excited state manifold⁷ and is not modelled. The smooth curve in the ns decay is a fit to the data using a sum of exponentials; these consist of a component that decays with the laser instrument response function, a 170-nm component, and a component with a decay time >10 μs . The 170-nm component yields the rate constant $k_{-1} = 6.0 \times 10^6 \text{ s}^{-1}$. The fits to the data were obtained using the Levenberg-Marquardt algorithm.

Table 1 Fluorescence quantum yields and decay times

Hairpin*	$\Phi_f^\dagger$	τ_s (ps) [‡]	τ_a (ns) [§]
T ₆ -A ₆	0.38 ± 0.02	2,000 (100)	
2G	0.018 ± 0.002	3.9 (23)	0.094 (77)
2G ₃	0.008 ± 0.001	3.9 (20)	0.10 (80)
3G	0.050 ± 0.005	31 (29)	1.9 (71)
3G ₃	0.050 ± 0.005	15 (35)	1.5 (50), 170 (15)
4G	0.20 ± 0.02	460 (40)	57 (60)
4G ₃	0.14 ± 0.01	610 (40)	220 (60)

* See Fig. 1 for structures.

[†] Quantum yield for fluorescence determined in deoxygenated 0.1 M aqueous NaCl solution using phenanthrene as the reference standard⁷.

[‡] Stilbene singlet decay time ($\tau_s = 1/k_{cs}$, except for T₆-A₆). Normalized decay amplitude in parentheses.

[§] Stilbene anion radical decay time ($\tau_a = 1/k_{cr}$, except for 3G₃ and 4G₃ as noted in text).

|| Data from ref. 7.

¶ No charge separation observed. Rate constant for stilbene singlet state decay.

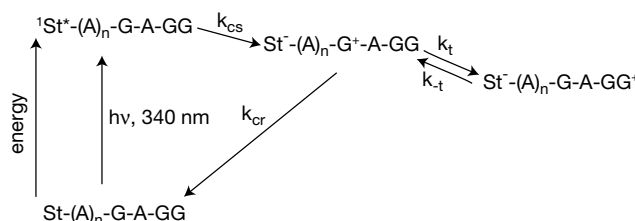


Figure 4 Kinetic scheme for processes occurring in a hairpin containing a single guanine proximal to the stilbene linker and a GG step distal from the stilbene linker. These

processes are charge separation (k_{cs}), charge recombination (k_{cr}), and hole transport (k_t and k_{-t}). Only the guanine-containing hairpin arm is shown.

ments at times > 5 ns utilized a laser system based on a frequency tripled Nd-YAG laser (355-nm excitation, 5-ns total instrument response function)^{6,7}. The initially formed stilbene singlet state $^1\text{St}^*$ decays to yield the anion radical $\text{St}^{\cdot-}$ with a decay time τ_s , and $\text{St}^{\cdot-}$ returns to the ground state with decay time τ_a . These decay processes can be fitted to a single exponential, except in the case of 3G_3 for which dual exponential decay of $\text{St}^{\cdot-}$ is observed (Fig. 3). Decay times are summarized in Table 1.

According to the kinetic scheme in Fig. 4, the singlet state decay time (τ_s) is determined by the rate constant for charge separation (k_{cs}) which should not be affected by the presence of a GG step remote from the stilbene linker⁷. In accord with this expectation, values for Φ_F , the $^1\text{St}^*$ fluorescence quantum yield and τ_s are either unchanged or changed only slightly for 2G_3 - 4G_3 versus 2G - 4G (Table 1). For 2G - 4G the anion radical decay time, τ_a , is determined by the rate constant for charge recombination, k_{cr} (ref. 7). The occurrence of hole transport from $\text{G}^{+\cdot}$ to GG would be expected to increase τ_a as a consequence of the longer distance between the anion and cation radicals and slower charge recombination for $\text{GG}^{+\cdot}$ versus $\text{G}^{+\cdot}$ (ref. 7). The value of τ_a for 2G_3 is similar to that for 2G , indicative of the failure of hole transport to compete with charge recombination in 2G_3 ($k_{cr} > k_t$). However, τ_a for 4G_3 is longer than that for 4G , indicative of the occurrence of efficient hole transport ($k_t > k_{cr}$). Thus the rate constant for hole transport is bracketed by the values of k_{cr} for 2G and 4G ($1 \times 10^{10} \text{ s}^{-1} > k_t > 2 \times 10^7 \text{ s}^{-1}$). In the case of 3G_3 two decay components are observed for the anion radical, one shorter and one longer than that for 3G (Table 1), indicative of the occurrence of both charge recombination and hole transport.

The long-lived decay components observed for 3G_3 (Fig. 3) can be analysed within the kinetic model in Fig. 4 to obtain k_t and k_{-t} . Using the two time constants for the biexponential decay of $\text{St}^{\cdot-}$, and their respective amplitudes, and the value of k_{cr} measured for 3G , the observed decay curve can be simulated using the exact analytical solution for the kinetic model in Fig. 4. This simulation yields $k_t = 5 \times 10^7 \text{ s}^{-1}$ and $k_{-t} = 6 \times 10^6 \text{ s}^{-1}$. Analysis of the $\text{St}^{\cdot-}$ decay for 4G_3 according to the kinetic model in Fig. 4 provides a value of $k_{-t} = 5 \times 10^6 \text{ s}^{-1}$, similar to the value of k_{-t} for 3G_3 . Charge recombination within the ion pairs $\text{St}^{\cdot-}-\text{A}_n-\text{G}-\text{A}-\text{GG}^{+\cdot}$ could, in principle, occur via a single-step superexchange process rather than a return hole transport mechanism. However, based on our studies of the distance dependence of charge recombination⁷, this should result in a large difference in $\text{St}^{\cdot-}-\text{A}_n-\text{G}-\text{A}-\text{GG}^{+\cdot}$ charge recombination rates rather than the small difference in the rates assigned to k_{-t} for 3G_3 versus 4G_3 .

The value of $k_t = 5 \times 10^7 \text{ s}^{-1}$ is too slow to compete with charge recombination within contact singlet radical ion pairs, which typically occurs with rate constants $> 10^9 \text{ s}^{-1}$ (refs 6,7,20). Thus in order for hole transport to occur, it is necessary to decrease the rate constant for charge recombination. This can be accomplished by separation of the radical ions (as seen in the case of 3G_3 and 4G_3), by the use of triplet acceptors (for which charge recombination is spin forbidden)^{10,13}, or by independent hole generation in the absence of a nearby electron^{11,12}. The ratio $k_t/k_{-t} \approx 8$ determined for 3G_3 is consistent with the observation that oxidative cleavage of DNA

occurs predominantly at GG (or GGG) steps²¹ but that GG steps do not totally inhibit further hole migration^{10,13}.

Rate constants for hole transport processes in DNA are expected to be dependent upon the relative energies of the initial and final states^{15–17}. For hole transport processes between two single guanines or two GG steps, the forward and return processes should have similar rate constants ($k_t \approx k_{-t}$). Rate constants for these processes should also be dependent upon the number of A:T base pairs separating the guanines and the location of the guanines either in the same strand or opposite strands. In the case of hole transport processes involving multiple hops, the efficiency is expected to be determined by the number of hops and the competition between hole transport and chemical reactions that lead to strand cleavage^{11,12,22}. □

Received 14 December 1999; accepted 16 May 2000.

- Becker, D. & Sevilla, M. D. Chemical consequences of radiation damage in DNA. *Adv. Radiat. Biol.* **17**, 121–180 (1993).
- Armitage, B. Photocleavage of nucleic acids. *Chem. Rev.* **98**, 1171–1200 (1998).
- Langenbacher, T. *et al.* Substrate and temperature dependence of DNA photolyase repair activity examined with ultrafast spectroscopy. *J. Am. Chem. Soc.* **119**, 10532–10536 (1998).
- Ratner, M. A. & Jortner, J. in *Molecular Electronics* (eds Jortner, J. & Ratner, M.) 5–72 (Blackwell, Oxford, 1997).
- Anderson, R. F. & Wright, G. A. Energetics and rate of electron transfer from base radical anions to electron-affinic intercalators in aqueous solution. *Phys. Chem. Chem. Phys.* **1**, 4827–4831 (1999).
- Lewis, F. D. *et al.* Distance-dependent electron transfer in DNA hairpins. *Science* **277**, 673–676 (1997).
- Lewis, F. D. *et al.* Dynamics of photoinduced charge separation and charge recombination in synthetic DNA hairpins with stilbenedicarboxamide linkers. *J. Am. Chem. Soc.* **122**, 2889–2902 (2000).
- Harriman, A. Electron tunneling in DNA. *Angew. Chem. Int. Edn Engl.* **38**, 945–949 (1999).
- Fukui, K. *et al.* Distance dependence of electron transfer in acridine-intercalated DNA. *J. Photochem. Photobiol. B* **50**, 18–27 (1999).
- Núñez, M. E., Hall, D. B. & Barton, J. A. Long-range oxidative damage to DNA: effects of distance and sequence. *Chem. Biol.* **6**, 85–97 (1999).
- Meggers, E., Michel-Beyerle, M. E. & Giese, B. Sequence dependent long range hole transport in DNA. *J. Am. Chem. Soc.* **120**, 12950–12955 (1998).
- Giese, B. *et al.* On the mechanism of long-range electron transfer through DNA. *Angew. Chem. Int. Edn Engl.* **38**, 996–998 (1999).
- Henderson, P. T., Jones, D., Hampikian, G., Kan, Y. & Schuster, G. Long-distance charge transport in duplex DNA: the phonon-assisted polaron-like hopping mechanism. *Proc. Natl Acad. Sci. USA* **96**, 8353–8358 (1999).
- Priyadarshy, S., Risser, S. M. & Beratan, D. N. DNA is not a molecular wire: protein-like electron-transfer predicted for an extended π -electron system. *J. Phys. Chem.* **100**, 17678–17682 (1996).
- Jortner, J., Bixon, M., Langenbacher, T. & Michel-Beyerle, M. E. Charge transfer and transport in DNA. *Proc. Natl Acad. Sci. USA* **95**, 12759–12765 (1998).
- Bixon, M. *et al.* Long-range charge hopping in DNA. *Proc. Natl Acad. Sci. USA* **96**, 11713–11716 (1999).
- Berlin, Y. A., Burin, A. L. & Ratner, M. On the long-range charge transfer in DNA. *J. Phys. Chem.* **104**, 443–445 (2000).
- Letsinger, R. L. & Wu, T. Use of a stilbenedicarboxamide bridge in stabilizing, monitoring, and photochemically altering folded conformations of oligonucleotides. *J. Am. Chem. Soc.* **117**, 7323–7328 (1995).
- Lewis, F. D. *et al.* Structure and photoinduced electron transfer in exceptionally stable synthetic DNA hairpins with stilbenediether linkers. *J. Am. Chem. Soc.* **121**, 9905–9906 (1999).
- Lewis, F. D., Liu, X., Miller, S. E. & Wasielewski, M. R. Electronic interactions between π -stacked DNA base pairs and diphenylacetylene-4,4'-dicarboxamide in hairpin DNA. *J. Am. Chem. Soc.* **121**, 9746–9747 (1999).
- Burrows, C. J. & Muller, J. G. Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* **98**, 1109–1151 (1998).
- Steenken, S. Electron transfer in DNA? Competition by ultra-fast proton-transfer? *Biol. Chem.* **378**, 1293–1297 (1997).

Acknowledgements

We thank R. L. Letsinger for discussions. This work was supported by the Division of Chemical Sciences, Office of Basic Energy Sciences, US Department of Energy.

Correspondence and requests for materials should be addressed to F.D.L. (e-mail: lewis@chem.nwu.edu).

Phosphate concentrations in lakes

Jeff J. Hudson^{*†}, William D. Taylor[‡] & David W. Schindler^{*}

^{*} Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

[‡] Department of Biology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Phosphate is an important nutrient that restricts microbial production in many freshwater^{1–3} and marine environments^{4–6}. The actual concentration of phosphate in phosphorus-limited waters is largely unknown because commonly used chemical and radiochemical techniques overestimate the concentration^{7,8}. Here, using a new steady-state radiobioassay to survey a diverse set of lakes, we report phosphate concentrations in lakes that are orders of magnitude lower than estimates made spectrophotometrically or with the frequently used Rigler radiobioassay. Our results, combined with those from the literature, indicate that microbes can achieve rapid turnover rates at picomolar nutrient concentrations. This occurs even though these concentrations are about two orders of magnitude below the level where phosphate uptake is estimated to be half the saturation level for the picoplankton community. Also, while phosphate concentration increased with the concentration of total phosphorus and soluble reactive phosphorus in the lakes we sampled, the proportion of phosphate in the total phosphorus pool decreased from oligotrophic to eutrophic lakes. Such information, as revealed by the phosphate assay that we use here, should allow us to address hypotheses concerning the concentration of phosphate available to planktonic microorganisms in aquatic systems.

Phosphate, that is, PO_4^{3-} , has been routinely measured for decades in aquatic environments⁹. However, such measurements are widely recognized to overestimate phosphate^{7,8,10} when phosphate is at low concentrations. For example, although recent modifications to the Rigler radiobioassay¹¹ produce relatively low estimates of phosphate compared to spectrophotometric techniques, this assay estimates phosphate to be the sum of phosphate and the apparent half-saturation constant for phosphate transport for the mixed community, and so these components cannot be separated¹⁰. Although there are many hypotheses concerning the effects of grazers, light, mixing, carbon sources and other factors on the availability of phosphorus to microorganisms, direct assessment of phosphate concentration is not possible. A new approach is clearly required if we wish to test hypotheses concerning phosphate concentration in aquatic ecosystems.

Illuminated surface waters where photosynthesis occurs are often isolated from underlying nutrient-rich waters by distinct thermoclines, and rely on rapid phosphorus recycling to sustain productivity. In such environments, inputs of phosphate from the atmosphere, from horizontal and vertical transport and from larger organisms (for example, fish) are small compared to planktonic regeneration¹². Under these conditions, the concentration of phosphate is usually too small to be estimated with chemical measurements, the turnover of phosphate is rapid and the flux of phosphate into (uptake) and out of (regeneration) the plankton must be approximately equal^{13,14}. This relationship can be used to solve for concentrations of phosphate.

We have employed a new method for estimating the regeneration of dissolved phosphorus¹⁵ (defined as that phosphorus which passes through a filter of 0.2- μm pore size). Most of the phosphorus regenerated by the plankton is phosphate^{16–19}, and the rest is likely to

be substrates for phosphatases²⁰. Although phosphate uptake cannot be directly measured, the uptake constant (k) for phosphate can be easily and accurately measured¹⁰. Therefore, with the regeneration rate and the uptake constant measured, we can solve for the concentration of phosphate (that is, phosphate uptake = $k \times [\text{PO}_4^{3-}]$ = regeneration rate). We refer to this measure of phosphate as a steady-state estimate.

Phosphate concentrations were determined for 56 lakes from three major physiographic regions of North America (Rocky Mountains, Interior Plains and Canadian Shield), spanning a nutrient gradient of 0.058–4.5 μM of total phosphorus (TP, which is the concentration of all phosphorus in the water, including dissolved phosphorus and phosphorus in plankton). In addition, we obtained concentrations of soluble reactive phosphorus (SRP) for 14 of these lakes from the literature. The colorimetric determination of SRP is still widely used as an estimate of phosphate concentration (for example, see APHA²¹) and provides a contrast with our steady-state estimates. We also compare the steady-state phosphate with Rigler bioassay determinations of phosphate in two Canadian Shield lakes.

Uptake constants (k) for phosphate were rapid, 0.02–1.1 min^{-1} , and indicate that phosphorus was limiting in all lakes^{10,22}. Dissolved phosphorus regeneration had a range of 210–26,000 pM h^{-1} . The steady-state estimates of phosphate concentration were between 27 and 16,800 pM (Fig. 1). This range narrowed to 27–885 pM when three lakes were removed from the data set. These lakes were identified as outliers in the regression analysis of phosphate on TP. Phosphate concentrations in these lakes were orders of magnitude greater than the 53 other lakes and are probably not a result of analytical error. Instead, we suspect that these lakes were not phosphorus-limited to the same extent as the other lakes, and that therefore phosphate concentrations were elevated due to low demand. The slow turnover times of phosphate in these lakes (range 31–64 min) compared to the other 53 lakes (range 1–10 min) supports this interpretation.

Phosphate concentration tended to increase with TP, but the proportion of phosphate declined with increasing TP (the slope of the relationship in Fig. 1 is 0.745). Our results illustrate that even at

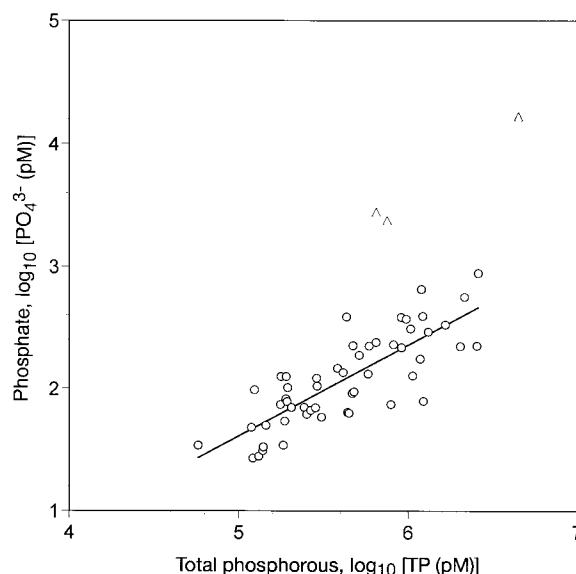


Figure 1 Steady-state phosphate estimates as a function of total phosphorus for 56 lakes. Phosphate concentration, $[\text{PO}_4^{3-}]$ in units of pM, tended to increase with total phosphorus (TP) concentration, $[\text{TP}]$ in units of pM: $n = 53$, $r^2 = 0.66$, $P < 0.0001$, $\log_{10}[\text{PO}_4^{3-} (\text{pM})] = 0.745(\log_{10}[\text{TP} (\text{pM})] - 2.11)$. Data points are shown as circles, except for estimates from three lakes, which are considered to be outliers (triangles) and were not included in the model I least-squares regression analysis.

[†] Present address: Dorset Environmental Sciences Centre & Trent University, PO Box 39, Bellwood Acres Road, Dorset, Ontario POA 1E0, Canada.

minute concentrations, rapid cycling of phosphate can support high rates of production. That is, the turnover rate of particulate phosphorus (phosphorus bound in plankton) is not a function of TP (ref. 12), implying that the productivity of plankton on a per unit biomass basis remains similar across a broad range in biomass.

Although our steady-state estimates of phosphate are not contemporaneous with measurements of SRP, the magnitude of the differences and the consistency of our phosphate estimates indicate that there is a large discrepancy between phosphate and SRP (Fig. 2). In no instance, even in eutrophic Halfmoon Lake (with a TP value of $2.6 \mu\text{M}$), did SRP concentrations approximate the steady-state concentrations. There are two well-documented explanations for this discrepancy. First, the SRP approach requires a water-filtration step to isolate the soluble phosphorus from the particulate phosphorus (phosphorus bound in plankton). During this step an error is introduced because a portion of the particulate phosphorus is released into the soluble pool, probably as a result of cell damage²³. Second, the reagents that are used to determine SRP acidify the filtrate and lead to the release of bound phosphate⁸. Both steps lead to overestimates of actual phosphate. In past studies, TP and SRP were routinely measured together. The TP measured in these studies may be substituted into the relationship in Fig. 1 to convert past SRP measurements to steady-state phosphate. However, this relationship would only provide an approximate measure.

Rigler bioassay estimates of phosphate in Mouse and Ranger Lakes were between 6 and 38 nM and represent concentrations that are approximately two orders of magnitude greater than the concurrent steady-state estimates (Fig. 3). There is an obvious explanation for this discrepancy. The Rigler bioassay uses the Michaelis-Menten equation and uptake velocities for different additions of phosphate to determine the sum of the half-saturation constant for uptake (K_s) of the added phosphate and the unknown ambient phosphate concentration²⁴. Therefore, the assay can only provide a potential upper limit of phosphate, because the phosphate concentration cannot be separated mathematically from the apparent K_s for the mixed community. Our results suggest that this apparent K_s is much greater than the phosphate concentration. Therefore, the Rigler bioassay is essentially an estimate of K_s and not of phosphate.

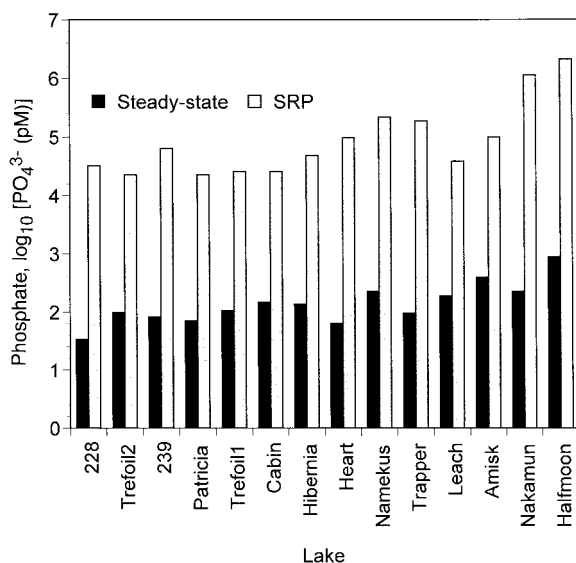


Figure 2 Comparison of steady-state and soluble reactive phosphorus estimates of phosphate concentration in a subset of 14 lakes. Lakes increase in nutrient content from the left (TP = $0.06 \mu\text{M}$) to the right (TP = $2.6 \mu\text{M}$). Soluble reactive phosphorus (SRP) concentrations are 2–3 orders of magnitude greater than steady-state concentrations. Summer SRP concentrations were obtained from the literature and government documents, and were not sampled concurrently with steady-state concentrations.

Despite these shortcomings, this assay has been used extensively for the past three decades to provide the lowest routine estimates of phosphate. Bentzen and Taylor¹⁰ provide a complete description of the assay.

One of the assumptions of the steady-state technique is that uptake of phosphate is equal to regeneration of phosphate. There are instances when this assumption may not be appropriate. When the turnover rate of phosphate by plankton is slow, indicating that the phosphate pool is large relative to the rate of phosphate uptake, regeneration and uptake may not be in balance. In addition, when the plankton community is increasing in biomass and drawing down the phosphate concentration, uptake would presumably exceed regeneration. More generally, our technique will underestimate phosphate in proportion to the relative importance of new versus recycled phosphate. Therefore, it should be applied only when the turnover rate of the phosphate is rapid relative to the supply from external sources. We do not know the extent of these errors in the current data set; however, we suspect that they are small because total planktonic biomass does not change rapidly relative to the turnover rate of phosphate during stratified conditions when phosphate concentration is low.

We report the smallest concentrations of phosphate for any aquatic system (that is, 27 pM). These concentrations are extremely low for a macronutrient, but equivalent concentrations exist for micronutrients. For example, dissolved Fe, Mn and Zn have been reported at picomolar concentrations in open-ocean surface waters^{25,26}. With a modified Rigler bioassay, phosphate concentrations have been measured in the 5 to 10 nM range in the phosphorus-limited Sargasso Sea⁴. If the modified Rigler bioassay consistently overestimates phosphate by two orders of magnitude, then these Sargasso Sea concentrations would be equivalent to 50 to 100 pM of steady-state phosphate. The low concentration and rapid turnover of phosphate supports the idea that biological processes (that is, uptake and regeneration) regulate limiting nutrient concentrations. Interestingly, the smallest concentrations reported for micro- and macronutrients appear to be in the low picomolar range. These concentrations may reflect the physiological limits to microbial scavenging of nutrients.

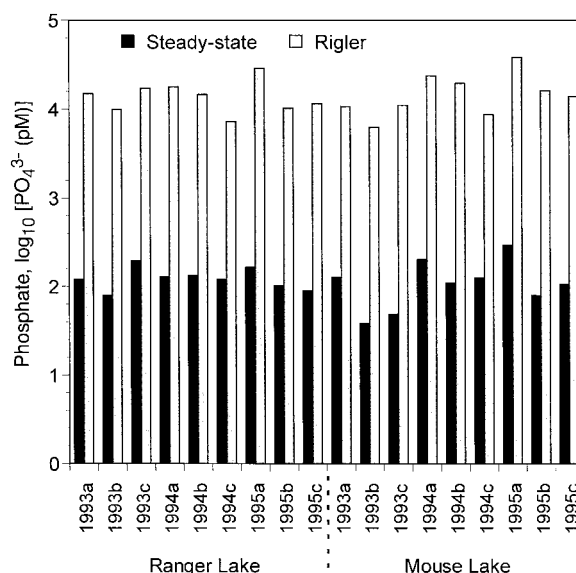


Figure 3 Comparison of steady-state and Rigler bioassay estimates of phosphate concentration in Mouse and Ranger Lakes. The comparison is based on concurrent measurements. Letters that follow each year on the x-axis refer to the time of sampling within a year: early summer (a), mid-summer (b) and late summer (c). Rigler bioassay concentrations are approximately two orders of magnitude greater than steady-state concentrations.

There are few studies that directly support the low estimates of phosphate that we have obtained with our steady-state bioassay; however, two single-lake studies made similar observations. Fisher and Lean⁷ measured phosphate concentrations by dialysis of radio-labelled lake water followed by column chromatography. With this elegant approach, concentrations of 500 pM were measured in mesotrophic Jacks Lake. In the same lake, a model of plankton interactions and a mass balance approach were used to deduce that the phosphate concentration must be about 300 pM (ref. 23). Dodds¹³ used a modified Michaelis–Menten model in combination with regeneration measurements by isotope dilution, and reported phosphate concentrations between 80–140 pM in oligotrophic Flathead Lake.

Although our steady-state bioassay is not suitable for routine monitoring, it is appropriate for testing hypotheses that address the phosphate concentration in phosphorus-limited pelagic systems. Our approach may also be adopted to examine the concentration of other nutrients. Most important, our results indicate that phosphate is at picomolar concentrations in lakes across a broad range in total phosphorus, and orders of magnitude lower than measured with the best widely used methods. □

Methods

Each lake was sampled once, except for Mouse and Ranger Lakes and three alpine lakes, which were sampled more frequently and over multiple years. Mean values are presented for these lakes. Only lakes that had a maximum depth equal to, or greater than 4 m were considered, in order to minimize benthic effects on the pelagic zone. We intended to sample only lakes with TP concentrations below 3.2 µM; however, one lake was sampled that exceeded this concentration (Islet). The water columns of 47 lakes were thermally stratified and the remaining lakes ($n = 9$) had isothermal water columns. Water (20 litres) was removed from a central location at the mid-epilimnetic depth of each lake with a Van Dorn sampler and placed in 20-litre polyethylene containers (acid washed) that were held in coolers. Isothermal lakes were sampled just below the surface (<1 m). Rigler bioassay¹⁰ determinations of phosphate were conducted on mid-epilimnetic water collected from a central station in Mouse and Ranger Lakes¹⁵. This water was collected at the same depth on the same day, or one day before the collection of water for the measurement of phosphorus regeneration¹⁵. Phosphate uptake constants¹⁰ and phosphorus regeneration rates¹⁵ were determined for all lakes. Standard chemical estimates of phosphate (SRP) were obtained from the literature for 14 of the study lakes (Fig. 2). SRP concentrations were provided by ref. 27 for Lake 239; ref. 28 for Heart, Namekus and Trapper Lakes; ref. 29 for Amisk and Nakamun Lakes; The Freshwater Institute (Winnipeg, Manitoba) for Lake 228; Environment Canada (Regina, Saskatchewan) for Trefoil 1 and 2, Patricia, Cabin, Hibernia, and Leach Lakes; and Alberta Environment (Edmonton, Alberta) for Halfmoon Lake.

Received 8 November 1999; accepted 18 May 2000.

- Schindler, D. W. Evolution of phosphorus limitation in lakes. *Science* **195**, 260–262 (1977).
- Wetzel, R. G. *Limnology* 2nd edn (Saunders, New York, 1983).
- Edmondson, W. T. *The Uses Of Ecology: Lake Washington And Beyond* 1st edn (Univ. Washington Press, Seattle, 1991).
- Cotner, J. B. et al. Phosphorus-limited bacterioplankton growth in the Sargasso Sea. *Aquat. Microb. Ecol.* **13**, 141–149 (1997).
- Karl, D. et al. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**, 533–538 (1997).
- Thingstad, T. F., Zweifel, U. L. & Rassoulzadegan, F. P. Limitation of heterotrophic bacteria and phytoplankton in the northwest Mediterranean. *Limnol. Oceanogr.* **43**, 88–94 (1998).
- Fisher, T. R. & Lean, D. R. S. Interpretation of radiophosphate dynamics in lake waters. *Can. J. Fish. Aquat. Sci.* **49**, 252–258 (1992).
- Baldwin, D. S. Reactive “organic” phosphorus revisited. *Wat. Res.* **32**, 2265–2270 (1998).
- Atkins, W. R. G. The phosphate content of fresh and salt waters in its relationship to the growth of the algal plankton. *J. Mar. Biol. Assoc.* **13**, 119–150 (1923).
- Bentzen, E. & Taylor, W. D. Estimating Michaelis–Menton parameters and lake water phosphate by the Rigler bioassay: importance of fitting technique, plankton size and substrate range. *Can. J. Fish. Aquat. Sci.* **48**, 73–83 (1991).
- Rigler, F. Radiobiological analysis of inorganic phosphorus in lakewater. *Verh. Int. Verein. Theor. Angew. Limnol.* **16**, 465–470 (1966).
- Hudson, J. J., Taylor, W. D. & Schindler, D. W. Planktonic nutrient regeneration and cycling efficiency in temperate lakes. *Nature* **400**, 659–661 (1999).
- Dodds, W. K. What controls levels of dissolved phosphate and ammonium in surface waters. *Aquat. Sci.* **55**, 132–142 (1993).
- Harrison, W. G. Nutrient recycling in production experiments. *ICES Mar. Sci. Symp.* **197**, 149–158 (1993).
- Hudson, J. J. & Taylor, W. D. Measuring regeneration of dissolved phosphorus in planktonic communities. *Limnol. Oceanogr.* **41**, 1560–1565 (1996).
- Peters, R. H. & Lean, D. R. S. The characterization of soluble phosphorus released by limnetic zooplankton. *Limnol. Oceanogr.* **18**, 270–279 (1973).
- Peters, R. H. & Rigler, F. H. Phosphorus release by *Daphnia*. *Limnol. Oceanogr.* **18**, 821–839 (1973).

- Ferrante, J. G. The characterization of phosphorus excretion products of a natural population of limnetic zooplankton. *Hydrobiologia* **50**, 11–15 (1976).
- Taylor, W. D. & Lean, D. R. S. Radiotracer experiments on phosphorus uptake and release by limnetic microzooplankton. *Can. J. Fish. Aquat. Sci.* **38**, 1316–1321 (1981).
- Bentzen, E. & Taylor, W. D. Estimating organic P utilization by freshwater plankton using (32P) ATP. *J. Plank. Res.* **13**, 1223–1238 (1991).
- American Public Health Association *Standard Methods for the Examination of Water and Wastewater* 19th edn (APHA, Washington, DC, 1995).
- Lean, D. R. S., Abbott, A. A. & Pick, F. R. Phosphorus deficiency of Lake Ontario plankton. *Can. J. Fish. Aquat. Sci.* **44**, 2069–2076 (1987).
- Taylor, W. D. & Lean, D. R. S. Phosphorus pool sizes and fluxes in the epilimnion of a mesotrophic lake. *Can. J. Fish. Aquat. Sci.* **48**, 1293–1301 (1991).
- Tarapchak, S. J. & Herche, L. R. Orthophosphate concentrations in lake water: analysis of Rigler’s radiobioassay method. *Can. J. Fish. Aquat. Sci.* **45**, 2230–2237 (1988).
- Bruland, K. W., Donat, J. R. & Hutchins, D. A. Interactive influences of bioactive trace metals on biological production in oceanic waters. *Limnol. Oceanogr.* **36**, 1555–1577 (1991).
- Landry, M. R. et al. Iron and grazing constraints on primary production in the central equatorial Pacific: An EqPac synthesis. *Limnol. Oceanogr.* **42**, 405–418 (1997).
- Schindler, D. W. in *Proc. IBP-UNESCO Symposium on Productivity Problems of Freshwaters* (eds Kajak, Z. & Hillbright-Ilkowska, A.) 303–331 (UNESCO, Warsaw, 1972).
- Canadian Park Service. *Prince Albert National Park Resource Description and Analysis*. Vol. 2, p. 4–40 (Environment Canada, Parks, Prairie and Northern Region, Winnipeg, Manitoba, 1986).
- Prepas, E. E. Orthophosphate turnover time in shallow productive lakes. *Can. J. Fish. Aquat. Sci.* **40**, 1412–1418 (1983).

Acknowledgements

We thank T. Paul, W. Bell, R. Hazewinkel, J. Almond, S. Leung, T. MacDonald, J. Huvane, B. Parker and N. McMaster for field and laboratory assistance. We also thank the Dorset Environmental Sciences Centre, D. McQueen and the Freshwater Institute for assistance, laboratory space, equipment and accommodation. D. Lean, P. Dillon, G. Mierle and K. Somers provided constructive criticism of the manuscript. A scholarship (NSERC) and a Killam Post-Doctoral Fellowship (University of Alberta) to J. J. H. and NSERC research grants to W. D. T. and D. W. S. supported this research.

Correspondence and requests for materials should be addressed to J. J. H. (e-mail: hudsonje@ene.gov.on.ca).

The sedimentary structure of linear sand dunes

C. S. Bristow*, S. D. Bailey* & N. Lancaster†

*School of Earth Sciences, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

†Desert Research Institute, UCCSN, 2215 Raggio Parkway, Reno, Nevada 89512, USA

Linear sand dunes—dunes that extend parallel to each other rather than in star-like or crescentic forms—are the most abundant type of desert sand dune¹. But because their development and their internal structure are poorly understood, they are rarely recognized in the rock record². Models of linear dune development^{2–6} have not been able to take into account the sub-surface structure of existing dunes, but have relied instead either on the extrapolation of short-term measurements of winds and sediment transport or on observations of near-surface internal sedimentary structures. From such studies, it has not been clear if linear dunes can migrate laterally^{2,7,8}. Here we present images produced by ground penetrating radar showing the three-dimensional sedimentary structure of a linear dune in the Namib sand sea, where some of the world’s largest linear dunes are situated. These profiles show clear evidence for lateral migration in a linear dune. Moreover, the migration of a sinuous crest-line along the dune produces divergent sets of cross-stratification, which can become stacked as the dune height increases, and large linear dunes can support superimposed dunes that produce stacked sets of trough cross-stratification. These clear structural signatures of linear dunes should facilitate their recognition in geological records.

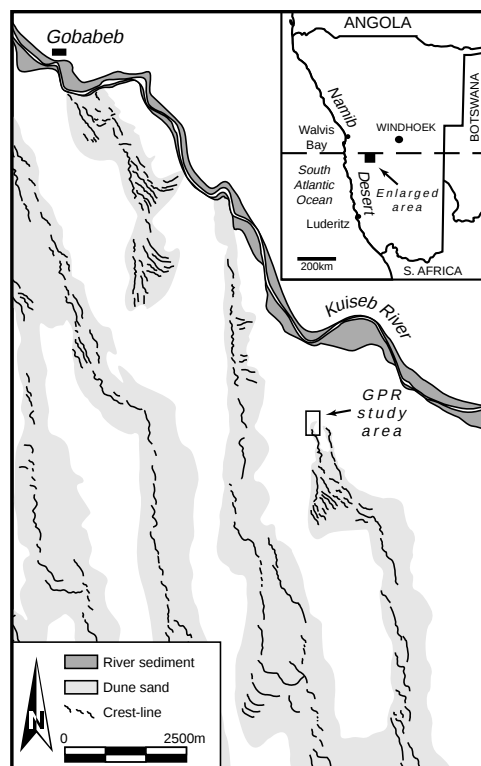


Figure 1 Locality map. The study site is at the northern end of a complex linear megadune in the Namib sand sea, Namibia.

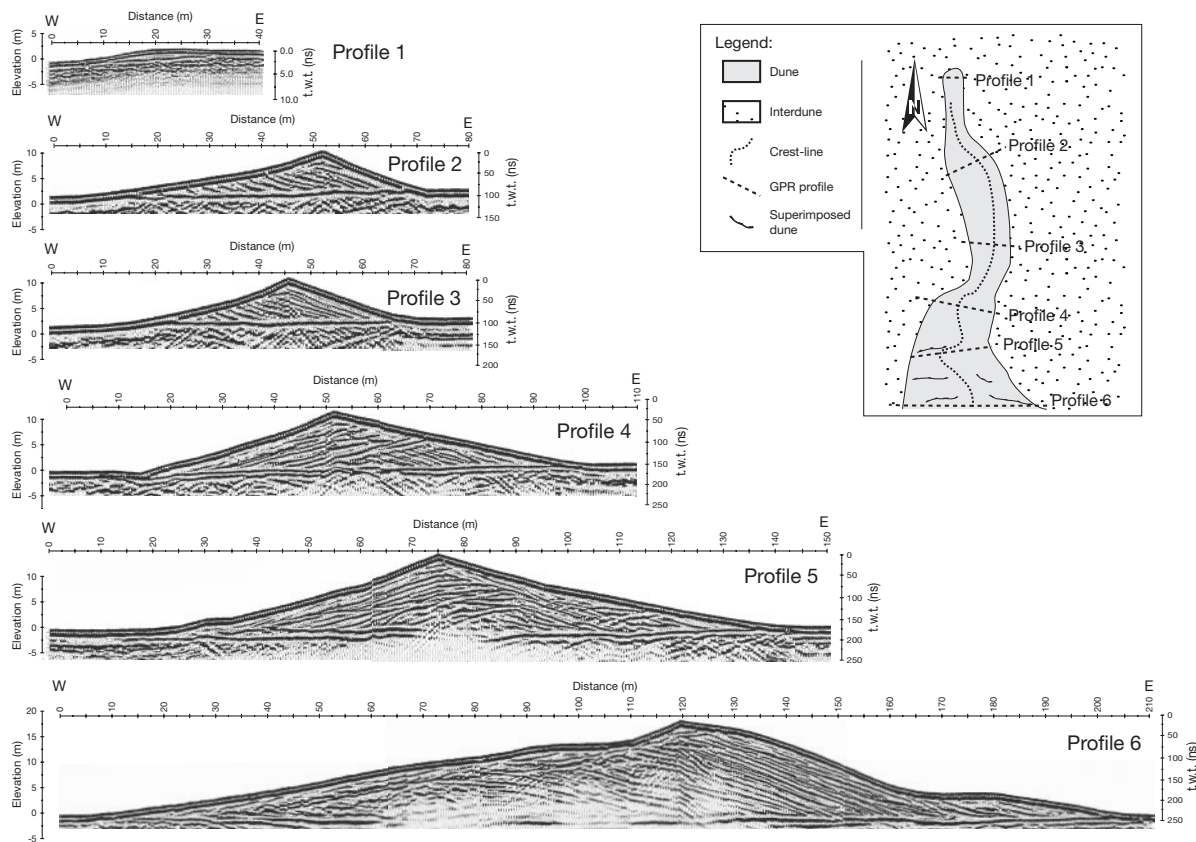


Figure 2 Dune structure from GPR profiles. GPR profiles across a sinuous linear dune show excellent resolution of sedimentary structures. On profile 1, structures are below the resolution of the GPR. Profiles 2 and 3 show planar sets of cross-stratification with dominant dips towards the east. Profiles 4 and 5 show bidirectional dips formed by northward migration of the sinuous crest-line. Profile 6 shows dominant dips to the east flanked by sets of trough cross-stratification from superimposed dunes. Profile 1 was

A topographic survey and ground penetrating radar (GPR) profiles were collected across and along the crest of the northern, leading edge of a sinuous linear dune located 15 km southeast of Gobabeb ($23^{\circ} 37' S$, $15^{\circ} 05' E$) on the northern edge of the Namib sand sea (Fig. 1). In this area there are marked seasonal variations in the wind regime. Winds from the southwest and the SSW are dominant during the summer months, while during the winter strong, but less frequent, winds blow from the east. The resultant sand drift direction is towards the NNE⁹. The northern end of the study dune is a low mound with no discernible slip face; to the south, the dune has a triangular cross-section with slopes (up to angle of repose on both sides) rising to a sharp crest. The crest-line is straight to slightly sinuous; sinuosity increases towards the south, with crest-line elevation rising to 20 m, and dune width increasing from 40 to 200 m. Superimposed crescentic dunes occur to the south where dune height exceeds 15 m. They are best developed where the sinuous dune flanks are convex in plan view.

We collected around 1.9 km of GPR profiles in a series of six profiles across the dune with one line along the dune crest. Data were collected with 100- and 200-MHz antennae. The 100-MHz antennae were spaced at 1 m with a step size of 0.5 m; 200-MHz antennae were spaced at 0.5 m with a step size of 0.2 m, except along the crest line where a step size of 0.5 m was used. A velocity of 0.17 m ns^{-1} has been calculated from a common midpoint survey, indicating theoretical values for resolution of 0.425 m for 100-MHz antennae and 0.21 m for 200-MHz antennae.

The GPR profiles show excellent resolution of sedimentary structures within the dune (Fig. 2). The sub-horizontal reflection at the base of each profile comes from the top of the underlying

collected using the higher-frequency 200-MHz antennae; the other lines were collected using 100-MHz antennae. t.w.t., two-way travel time. Topography was measured in the field at 5-m intervals and at breaks of slope along the profiles using a total station (an instrument that combines electronic distance measurement with a theodolite). Topographic correction has been applied to each profile based on an arbitrary local datum.

surface, because there is a strong contrast in dielectric properties between the modern dune sand and the underlying carbonate-cemented Tertiary fluvial gravel and aeolian sandstone. Reflections within the dune come from sets of cross-stratification¹⁰. The low northern end of the dune (profile 1 in Fig. 2) shows no discernible structure because it is composed of wind-ripple laminae which are below the resolution of the GPR; this sand has been blown north from the main dune. The easterly dipping sets of cross-stratification in profiles 2 and 3 (Fig. 2) were formed during northeastward movement of the dune towards the resultant drift direction. A profile along the dune crest (not illustrated) shows sub-horizontal planar reflectors dipping very gently towards the north, confirming the planar tabular nature of the cross-stratification.

The bimodal dips in profiles 4 and 5 (Fig. 2) are attributed to northward extension of the sinuous dune, which results in vertically stacked sets of cross-stratification dipping to east and west, separated by erosional bounding surfaces. The lower easterly dips are consistent with the planar cross-strata in the straighter, more northerly, parts of the dune (profiles 2 and 3). Truncation of the lower reflectors occurs within the concave section of the dune sinuosity. Younger sets that downlap onto the erosion surface are formed by westward deflection of the crest line as a convex sinuous element moves north. Repeated westward and eastward shifts in position of the crest-line due to the northward translation of crest-line sinuosity results in offset vertical stacking of sets of cross-stratification (profile 5). The easterly dips within profile 6 indicate dune migration towards the east, towards the resultant drift direction. Sets of trough cross-stratification on profiles 5 and 6 (Fig. 2) were formed by superimposed dunes migrating north along the dune flanks.

We recognize five stages in the evolution of a linear dune (Fig. 3): (1) an initial extension of wind-rippled sand; (2) development of a relatively straight-crested transverse element that migrates in response to the dominant wind direction, producing planar sets of cross-stratification; (3) as the dune grows, sinuosity in the dune crest-line develops, probably as a result of some initial irregularity in morphology. The cross-sectional area also increases exponentially to the point where it cannot be remodelled in each wind season and form-flow interactions become significant. The dune develops a morphology that is controlled by more than one wind direction, and the dune's sedimentary structure changes, with the formation of two opposing sets of cross-stratification (Fig. 3). The bimodal

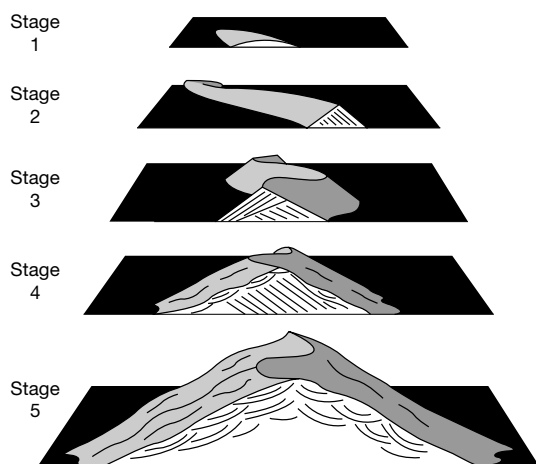


Figure 3 The five-stage model of linear dune evolution. Stage 1, initial plume of wind rippled sand. Stage 2, slip-face development and migration towards the resultant drift direction produces simple sets of cross-stratification resembling a transverse dune. Stage 3, lateral displacement of the sinuous crest-line creates bidirectional dips. Stage 4, flow deflection along dune flanks leads to the development of superimposed dunes. Stage 5, complex linear megadunes are dominated by sets of trough cross-stratification from superimposed dunes.

dips are attributed to stacking of sinuous elements migrating along the dune. This contrasts with the reversing crest model³, and at this stage the structure is similar to the model of Tsoar⁶; (4) superimposed dunes appear with increasing dune height, width and sinuosity, because sand transport rates are then sufficient to support migrating dunes on the dune slopes¹¹. The dominant southerly and southwesterly winds produce small superimposed dunes with slip faces (1–2 m high) facing north; such slip faces are best developed on the outer, convex slopes of sinuous dune elements, where winds are deflected along both windward and lee-side dune flanks (Fig. 4). The development of superimposed bedforms results in another change in the style of deposition, with sets of trough cross-stratification overlying larger sets of cross-stratification with a dominant dip towards the resultant drift direction which resembles model C of Rubin and Hunter². Profile 6 (Fig. 2) also shows some similarity to the supposed internal structure of the 'whalebacks' of the Egyptian sand sea, as given by Bagnold in ref. 3. However, Bagnold attributed lateral migration to a change in wind regime rather than a response to sediment transport within a bimodal wind regime. (5) Sets of cross-stratification from superimposed bedforms then start to dominate dune sedimentary structure (Fig. 3). GPR profiles across larger complex linear megadunes in the same area are dominated by sets of trough cross-stratification from superimposed dunes resembling model E of Rubin and Hunter². The model of Bagnold³ is not supported by our observations, despite the bidirectional winds and observations of reversing crest-lines in this area⁷.

Dune topography has a strong effect on wind flow lines and sediment transport over the dune¹². Idealized wind flow lines for southwesterly and easterly winds based on field observations are shown in Fig. 4. On windward slopes, concave surfaces act to focus flow lines, locally increasing sediment transport at low points in the

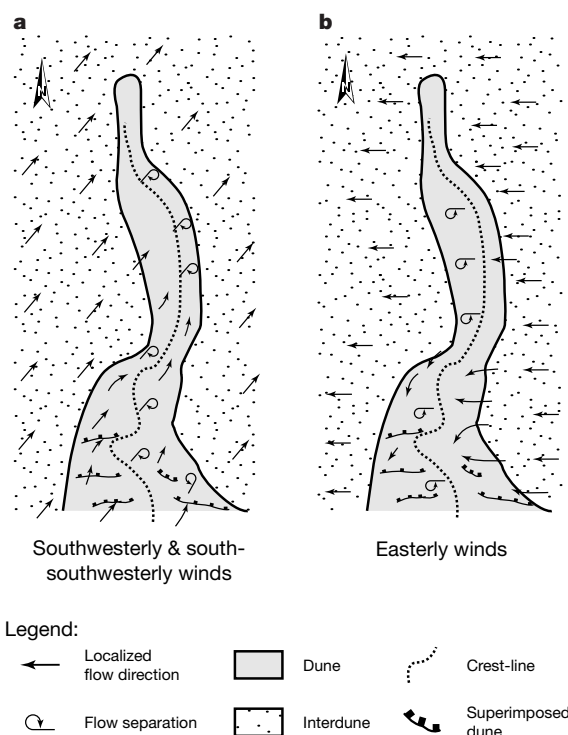


Figure 4 Wind flow lines. Idealized wind flow lines over a sinuous linear dune with a bidirectional wind regime, based on a field sketch of the study dune. Wind is locally focused where the windward dune flank is concave and deflected where the dune flank is convex. Flow separation occurs where the wind is almost perpendicular to the crest-line. Flow lines are locally deflected along both lee and windward slopes. Wind flow along the dune flanks promotes the development of superimposed dunes which are often best developed where the dune flanks are convex, especially on the eastern slope which is the lee side with respect to southwesterly winds.

crest-line. This results in increased deflection and sinuosity of the crest-line. Where the windward slope is convex in plan, wind flow lines tend to diverge and may be deflected along the face, locally increasing transport along the dune. On the lee side of the dune flow separation occurs where the wind is perpendicular to the crest-line; but separation does not always extend down or along the whole lee-side slope¹². Lee-side flows may be deflected along the lee-side slope, especially on the convex dune slopes. Wind deflection over convex surfaces on both windward and lee slopes increases sediment transport along the dune and promotes the development of superimposed dunes. The oblique orientation of the dunes with respect to winds from the southwest and SSW leads to deflection along both dune flanks. Superimposed dunes are often best developed on the eastern side of complex linear dunes in the Namib sand sea, which is the lee side with respect to the more common SSW and south-westerly winds. The oblique orientation to the resultant transport vector may indicate alignment normal to the maximum gross transport¹³.

By combining geomorphic and geophysical investigations into the morphology and structure of a linear dune, we have been able to determine its development and evolution. The high resistivity of aeolian sands gives good penetration (15–20 m), and the large sedimentary structures within dune sands are clearly imaged. The ability of GPR to image deep within the dune has been proven, and this technique provides a unique insight into the internal structure of dunes that cannot be achieved by any other non-destructive or geophysical method. Changes in dune morphology are reflected in the internal structures, and the results of this survey show that with increasing size and morphologic complexity, linear dunes become dominated by sets of cross-stratification from superimposed dunes. A simple, straight-crested linear dune which has migrated laterally could be mistaken for a transverse dune. The development of superimposed dunes results in the formation of stacked sets of trough cross-stratification that dominate the sedimentary structures of large, complex, linear megadunes, as well as many ancient aeolian sandstones, suggesting that deposits formed by large linear dunes may be much more common in the rock record than was previously thought. □

Received 29 October 1999; accepted 11 May 2000.

1. Lancaster, N. Linear dunes. *Prog. Phys. Geogr.* **6**, 475–504 (1982).
2. Rubin, D. M. & Hunter, R. E. Why deposits of longitudinal dunes are rarely recognised in the geologic record. *Sedimentology* **32**, 147–157 (1985).
3. Bagnold, R. A. *The Physics of Blown Sand and Desert Dunes* (Methuen, London, 1941).
4. McKee, E. & Tibbitts, G. C. Jr Primary structures of a seif dune and associated deposits in Libya. *J. Sedim. Petrol.* **34**, 5–17 (1964).
5. McKee, E. Sedimentary structures in dunes of the Namib desert, South West Africa. *Geol. Soc. Am. Spec. Pap.* **188**, 64 (1982).
6. Tsoar, H. Internal structure and surface geometry of longitudinal seif dunes. *J. Sedim. Petrol.* **52**, 823–831 (1982).
7. Hesp, P., Hyde, R., Hesp, V. & Zengyu, Q. Longitudinal dunes can move sideways. *Earth Surf. Processes Landforms* **14**, 447–451 (1989).
8. Rubin, D. M. Lateral migration of linear dunes in the Strzelecki Desert, Australia. *Earth Surf. Processes Landforms* **15**, 1–14 (1990).
9. Livingstone, I. Monitoring surface change on a Namib linear dune. *Earth Surf. Processes Landforms* **14**, 317–332 (1989).
10. Bristow, C., Pugh, J. & Goodall, T. Internal structure of aeolian dunes in Abu Dhabi determined using ground penetrating radar. *Sedimentology* **43**, 995–1003 (1996).
11. Lancaster, N. *Geomorphology of Desert Dunes* (Routledge, London, 1995).
12. Tsoar, H. Dynamic processes acting on a longitudinal (seif) sand dune. *Sedimentology* **30**, 567–578 (1983).
13. Rubin, D. M. & Hunter, R. E. Bedform alignment in directionally varying flows. *Science* **237**, 276–278 (1987).

Acknowledgements

We thank the staff of DRFN at Gobabeb for their assistance and logistical support without which this project would not have been possible; the Namibian Ministry of Environment and Tourism for permission to undertake research within the Namib Naukluft National Park; and D. Upton for his assistance in the field and for drafting the figures. This work was supported by the NERC; S.D.B. thanks the NERC for a studentship.

Correspondence and requests for materials should be addressed to C. S. B. (e-mail: c.bristow@bbk.ac.uk).

The formation of plagioclase chains during convective transfer in basaltic magma

Anthony R. Philpotts & Loretta D. Dickson

Department of Geology and Geophysics, University of Connecticut, Storrs, Connecticut 06269, USA

The basaltic rock in the lower part of the thick Holyoke lava flow in Connecticut and Massachusetts has been shown to have a remarkable texture, with crystals of feldspar linked together in a continuous three-dimensional network of chains¹. Heating experiments have revealed that this network persists to temperatures where the rock is 75% liquid², and therefore the network was interpreted to have formed at an early stage of crystallization and to have played an important role in the compaction of crystal mush in the lower part of the flow^{3,4}. Despite the texture's importance to our understanding of how such basalt flows form, the origin of the texture has remained uncertain^{1–3}. Here we show that, although the network is present in the lower third of the flow, it was actually formed in the upper solidification front and was transported down in plumes of dense crystal mush. Convection of this type has been postulated for intrusive magma chambers, but corroborative field evidence has been equivocal, especially in lava lakes and flows. Preservation of the roof-generated texture in the lower part of a thick flood-basalt flow therefore constitutes important evidence for the role of convection in the solidification and differentiation of a simple magma sheet.

The Mesozoic Holyoke flood basalt of Connecticut and Massachusetts forms a laterally extensive (125 km north–south, and 50 km east–west), thick (50–200 m) flow preserved in the down-faulted Hartford, Deerfield and Pomperaug basins. As in other thick flood-basalt flows, fractures propagating down from the top and up from the bottom generated the well-known entablature and colonnade structure. The boundary between these two opposing fracture sets marks approximately where the crystallization front moving down from the roof of the magma sheet met the front solidifying up from the base. Because cooling is more rapid from above, this boundary is located typically at one-third the height of a flow and may be as little as one-fifth if cooling is accelerated by water on the surface of the flow^{5–7}. In thick flood-basalt flows and lava lakes, however, this boundary is well above the one-third level and may actually be above the centre of the flow⁶. For example, where the Holyoke flow is only 50 m thick, the colonnade constitutes only 25% of the flow thickness, but where it is 200 m thick, it constitutes 62%. This upward displacement of the final solidification zone requires that material be transferred from the roof to the floor of the crystallizing sheet. Moreover, at least some of the solids must survive this transport through the central part of the sheet without being totally resorbed; that is, solids must arrive at the floor to displace the final liquid upward.

The textures of the entablature and colonnade in the Holyoke basalt, like those in many thick flood-basalt flows, are quite distinct and point to different modes of crystallization⁵. The entablature is characterized by an intersertal texture, defined by ophitically intergrown clusters of pyroxene and plagioclase (Fig. 1) separated by patches of mesostasis containing dendritic magnetite crystals and immiscible iron-rich liquid droplets (now glass) in a felsic matrix. Plagioclase crystals are elongate and commonly shaped like tuning forks, and even 80 m below the surface of the flow they still have aspect ratios as high as 70:1. The colonnade, in contrast, has an

intergranular texture with clusters of granular pyroxene grains separated by plagioclase crystals that are linked together in chains that form a three-dimensional network (Fig 1). The plagioclase crystals have a much smaller aspect ratio than do those in the entablature. Magnetite forms subhedral, equant crystals, and the residual liquid crystallized to granophyric intergrowths of tridymite and alkali feldspar.

The characteristics of the plagioclase-chain network have been revealed in considerable detail by serial sections of a number of samples and by X-ray computed tomography scans of one partly melted sample³. We summarize here only those features relevant to a discussion of the network's formation. The plagioclase chains consist of groundmass crystals, which are approximately 0.5 mm long and 0.1 mm wide. The chains are several crystals wide. They branch every 1 to 2 mm and typically thicken at junctions. The rock contains about 5% plagioclase phenocrysts, which commonly occur at these junctions. Crystals in the chains are normally zoned from cores of An_{65} to An_{32} , but the maximum anorthite content at the contacts between individual crystals is An_{59} . (Here An_x indicates $x\%$ anorthite.) The crystals therefore grew separately for a period before becoming linked together. The patches of granophyre, which formed from residual liquid, occur only in contact with plagioclase chains and never with pyroxene grains. The 'mesh' size of the network coarsens upward in the flow, even though the grain size of the minerals does not change significantly with height.

The pyroxene crystals in the entablature and colonnade are distinct and indicate different modes of crystallization. All pyroxene in the entablature is intergrown with plagioclase in ophitic clusters. Thin plagioclase laths radiate from the centre of each cluster (Fig 1)

and become progressively more sodic as the pyroxene becomes more iron-rich toward the rim. From the core of a cluster, augite ($Wo_{30}En_{56}Fs_{14}$) and pigeonite ($Wo_5En_{78}Fs_{17}$) zone outward continuously through the field of subcalcic augite to compositions near $Wo_{33}En_{36}Fs_{31}$ (Wo, wollastonite; En, enstatite; Fs, ferrosilite). Small olivine crystals (now altered to clay minerals) occur in the central parts of many clusters where they are rimmed by the pigeonite. Figure 2 shows a pyroxene quadrilateral plot of analyses along radial traverses in a single ophitic cluster. The sample is from 79 m beneath the surface of the 200-m-thick flow, near the base of the entablature. Although the pyroxenes exhibit a systematic enrichment in iron toward the rim of the cluster, they completely disregard the equilibrium pyroxene solvus⁸. Such compositions are indicative of rapid crystallization, as for example, in supercooled lunar basalts⁹.

In the colonnade, pyroxenes occur as clusters of small grains, most exhibiting 120° triple junctions that are indicative of recrystallization (Fig. 1). Both augite and pigeonite are present and intimately mixed. Although both cover a wide range of iron/magnesium ratios, they have restricted calcium contents. Figure 2 shows a plot of these pyroxenes from a sample 43 m above the base of the flow. The augite and pigeonite plot on either side of the pyroxene solvus with compositions consistent with equilibration at magmatic temperatures between $1,170^\circ\text{C}$ and $1,000^\circ\text{C}$ (ref. 8). These pyroxenes must have cooled considerably more slowly than those in the entablature, despite the fact that the entablature sample was twice as far from a contact as the colonnade sample.

These textural and mineralogical differences between the entablature and colonnade indicate that the entablature crystallized far more rapidly than did the colonnade, and hence the

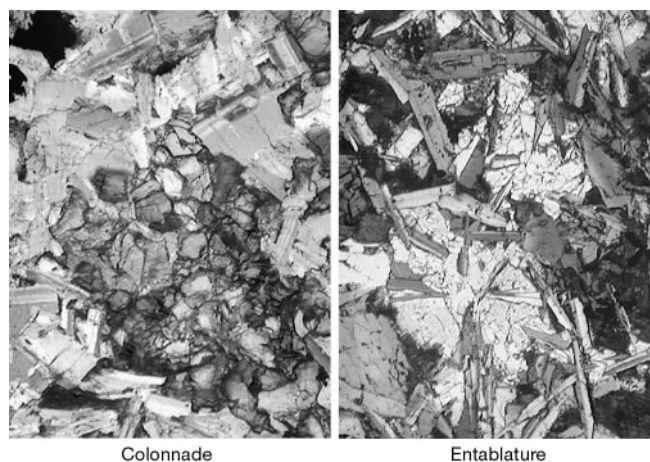


Figure 1 Photomicrographs showing the contrasting mode of occurrence of pyroxene in the entablature and colonnade of the Holyoke basalt. Both fields are 1 mm high, and the images were obtained using partly crossed polars. In the entablature, ophitic clusters of strongly zoned subcalcic augite and plagioclase are surrounded by earlier-crystallizing intersertal plagioclase crystals in a dark mesostasis. In the colonnade, clusters of

extremely fine granular augite and pigeonite are surrounded by plagioclase crystals that are linked together to form a three-dimensional network. The colonnade texture is interpreted to have formed through recrystallization of the entablature texture in the slower-cooled lower part of the flow.

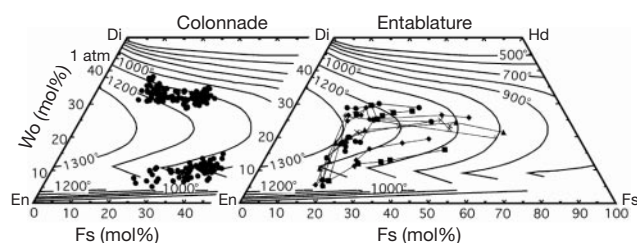


Figure 2 Analyses of pyroxene crystals in the Holyoke flow plotted in the pyroxene quadrilateral, showing the augite–pigeonite solvus⁸. Crystals from the entablature and colonnade were analysed by electron microprobe. The entablature analyses are from a single ophitic pyroxene cluster. The lines join successive analyses along radial traverses.

Most analyses fall within the pyroxene solvus. The colonnade analyses, which are all from one sample, plot on either side of the solvus at normal magmatic temperatures. Di, diopside; Hd, hedenbergite.

colonnade–entablature boundary should have been displaced downward. Instead, it occurs 120 m above the base of the 200-m-thick section. This can only mean that during the rapid crystallization in the upper solidification front, dense crystal mush sank to the floor of the magma sheet and thus displaced the final solidification zone upward⁶. Moreover, the colonnade–entablature boundary in this 200-m-thick section of the flow is continuously exposed for 4.6 km in the east face of the North Branford quarry of Tilcon Connecticut, Inc. Over this distance, the boundary remains at a constant height in the flow, indicating that the amount of material transferred from roof to floor is everywhere the same. This rules out the possibility that transfer resulted from accidental foundering of parts of the roof, which would have resulted in the boundary occurring at variable heights. The transfer of material from roof to floor is therefore considered to be an integral part of the solidification process in the upper crystallization zone.

The initial magma density, calculated using the MELTS software package¹⁰, was 2,670 kg m⁻³, but as plagioclase and pyroxene crystallized, its density steadily increased, reaching 2,690 kg m⁻³ after 33% crystallization. The plagioclase, with a density of 2,642 kg m⁻³, was slightly buoyant in this liquid, but the pyroxene, with a density of 3,300 kg m⁻³, gave the solids and residual liquid a bulk density of 2,750 kg m⁻³. The crystal mush in the upper solidification zone was therefore always gravitationally unstable. During the solidification of the upper 25 m of the flow, however, the mush was buoyed up by vesicles that rose from the flow's interior. The grain size in this vesicular zone increases with depth, as would be expected from a thickening solidification zone. Below 25 m, however, vesicles are absent and crystal mush in the lower part of the solidification zone would have been able to slough off and thus prevent the zone from continuing to thicken. The grain size consequently increases very little between 25 m and the bottom of the entablature at 80 m. It is worth noting that the flow does not contain multiple zones of vesicles, which in other thick flows have been shown to result from multiple pulses of magma inflating the flow¹¹. Their absence in the Holyoke basalt indicates that the flow was formed either primarily by one eruptive pulse or that multiple zones were later destroyed by convective processes in the ponded flow.

The texture of the material that sank from the inner part of the upper solidification zone would have resembled the material that remained to form the entablature, except for a lower degree of crystallinity. The texture in the colonnade can therefore be interpreted as a recrystallized modification of the entablature's texture. Once this connection is made, the plagioclase-chain network in the colonnade has a simple explanation.

In the entablature, early crystallizing groundmass plagioclase crystals (An_{65–59}) and phenocrysts were pushed aside by the growth of the later crystallizing ophitic pyroxene–plagioclase clusters to form an incipient plagioclase-chain network (Fig. 1). Residual liquid, which also concentrated between these pyroxene–plagioclase clusters, solidified in the entablature to form patches of mesostasis. When transported into the more slowly cooled floor zone, however, this intersertal liquid crystallized more slowly and, following fractionation, eventually produced the granophyric intergrowth that occurs exclusively in contact with plagioclase chains. On sinking to the floor, the pyroxene crystals in the ophitic clusters, which had initially grown rapidly with disequilibrium compositions, recrystallized into fine granular aggregates of augite and pigeonite with equilibrium compositions. This recrystallization was facilitated by the presence of melt, as indicated by the compositions of the pyroxenes on the solvus (Fig. 2). Most of the plagioclase crystals that were initially ophitically intergrown with the pyroxene were expelled from the clusters during the recrystallization and added to the surrounding intersertal crystals to widen the plagioclase chains forming the network. One of the puzzling features of the network when it was first described³ was that its 'mesh' size bore no apparent relation to the size of the fine granular pyroxene crystals

that filled the spaces between the chains. This mesh size is now interpreted as being determined by the size and spacing of the ophitic pyroxene–plagioclase clusters in the roof zone. The coarsening of the network away from the base of the flow with no corresponding coarsening of the individual crystals³ would be a natural consequence of the increasing size of the ophitic clusters away from the roof of the magma sheet. This steady increase in the network's mesh size is further evidence that the transfer of material from roof to floor was a continuous process. If foundering of roof material was a one-time event, only the foundered block would exhibit the recrystallized roof texture and the magma surrounding the sunken block would crystallize with a different texture. No evidence for such textural variability has been found in the lower part of the flow, despite detailed sampling on a meter scale. And finally, because the pyroxene–plagioclase clusters in the entablature are roughly spherical, the incipient plagioclase chain-network is isotropic. However, on the floor of the magma sheet, when the pyroxene crystals recrystallize, compaction is able to distort the network and leave a record of the amount of compaction³.

The upward displacement of the colonnade–entablature boundary in a flow that underwent rapid cooling from above is clear evidence for convective transfer of dense crystal mush from the roof to the floor of the magma sheet. Moreover, the preservation in the lower part of the flow of the roof-generated texture, albeit slightly modified by recrystallization, attests to the coherent nature of the mush as it sank to the floor. Had the mush not remained intact (with its bulk negative buoyancy), the slightly buoyant plagioclase crystals would have been left behind as the dense pyroxene and residual liquid sank. This would have led to chemical fractionation of the lava, which detailed analyses through the flow indicate did not occur⁴. The plagioclase-chain network, which had its origin in the upper solidification zone, therefore played an important role in binding the mush together and permitting otherwise buoyant plagioclase crystals to be transferred downward.

At least in the case of this thick flood-basalt flow, the simplest of all possible magma bodies, the plagioclase-chain network provides a simple solution to the long-standing petrologic problem of how plagioclase crystals, which are neutrally buoyant in most basaltic magma, can accumulate on the floor of layered mafic intrusions. □

Received 23 December 1999; accepted 12 May 2000.

- Philpotts, A. R., Shi, J. & Brustman, C. Role of plagioclase crystal chains in the differentiation of partly crystallized basaltic magma. *Nature* **395**, 343–346 (1998).
- Philpotts, A. R. & Carroll, M. Physical properties of partly melted tholeiitic basalt. *Geology* **24**, 1029–1032 (1996).
- Philpotts, A. R., Brustman, C., Shi, J., Carlson, W. D. & Denison, C. Plagioclase-chain networks in slowly cooled basaltic magma. *Am. Mineral.* **84**, 1819–1829 (1999).
- Philpotts, A. R., Carroll, M. & Hill, J. M. Crystal-mush compaction and the origin of pegmatitic segregation sheets in a thick flood-basalt flow in the Mesozoic Hartford basin, Connecticut. *J. Petrol.* **37**, 811–836 (1996).
- Long, P. E. & Wood, B. J. Structures, textures, and cooling histories of Columbia River basalt flows. *Geol. Soc. Am. Bull.* **97**, 1144–1155 (1986).
- Marsh, B. D. Crystal capture, sorting, and retention in convecting magma. *Geol. Soc. Am. Bull.* **100**, 1720–1737 (1988).
- DeGraff, J. M. & Aydin, A. Surface morphology of columnar joints and its significance to mechanics and direction of joint growth. *Geol. Soc. Am. Bull.* **99**, 605–617 (1987).
- Lindsley, D. H. Pyroxene thermometry. *Am. Mineral.* **68**, 477–493 (1983).
- Dowty, E., Keil, K. & Prinz, M. Lunar pyroxene-phyrical basalts: Crystallization under supercooled conditions. *J. Petrol.* **15**, 419–453 (1974).
- Ghiorsio, M. S. & Sack, R. O. Chemical mass transfer in magmatic processes. IV. A revised and internally consistent thermodynamic model for the interpolation and extrapolation of liquid-solid equilibria in magmatic systems at elevated temperatures and pressures. *Contrib. Mineral. Petrol.* **119**, 197–212 (1995).
- Self, S. et al. A new model for the emplacement of Columbia River basalts as large inflated pahoehoe lava flow fields. *Geophys. Res. Lett.* **23**, 2689–2692 (1996).

Acknowledgements

We thank F. T. Lane for permission to collect samples in the North Branford quarry, and B. Marsh for comments on the manuscript. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to A.R.P. (email: philpotts@geol.uconn.edu).

Valuation of consumption and sale of forest goods from a Central American rain forest

Ricardo Godoy*, David Wilkie†, Han Overman‡, Adoni Cubas§, Glenda Cubas||, Josefien Demmer‡, Kendra McSweeney¶ & Nicholas Brokaw#

* Department of Anthropology, Brandeis University, Waltham, Massachusetts 02254, USA

† Associates in Forest Research and Development, 18 Clark Lane, Waltham, Massachusetts 02154, USA

§ Honduras Coral Reef Fund, Apartado Postal 1053, La Ceiba, Atlántida, Honduras

|| Universidad Pedagógica Nacional, Apartado 3394, Miraflores Blvd, Tegucigalpa, M.D.C., Honduras

‡ Amsterdam Research Institute for Global Issues and Development Studies, University of Amsterdam, 1018 VZ Amsterdam, The Netherlands

¶ Department of Geography, McGill University, Montreal, Quebec, H3A 2K6 Canada

Manomet Center for Conservation Sciences, Box 1770, Manomet, Massachusetts 02345, USA

Researchers recognize that society needs accurate and comprehensive estimates of the economic value of rain forests to assess conservation and management options^{1–7}. Valuation of forests can help us to decide whether to implement policies that reconcile the value different groups attach to forests. Here we have measured the value of the rain forest to local populations by monitoring the foods, construction and craft materials, and medicines consumed or sold from the forest by 32 Indian households in two villages in Honduras over 2.5 years. We have directly measured the detailed, comprehensive consumption patterns of rain forest products by an indigenous population and the value of that consumption in local markets^{8,9}. The combined value of consumption and sale of forest goods ranged from US\$17.79 to US\$23.72 per hectare per year, at the lower end of previous estimates (between US\$49 and US\$1,089 (mean US\$347) per hectare per year)⁴. Although outsiders value the rain forest for its high-use and non-use values¹⁰, local people receive a small share of the total value. Unless rural people are paid for the non-local values of rain forests, they may be easily persuaded to deforest.

The Tawahka (population 1,000), a farming and foraging group of Amerindians, live in five villages along the Patuca River in a region of tropical moist forest of eastern Honduras. Research took place in two villages 16 km apart, Krausirpe and Yapuwas, that have different degrees of exposure to the market. Krausirpe is 30 km from the nearest market town and more affluent than Yapuwas, a more isolated village¹¹. The habitat, which extends 3 km upstream and 3 km downstream of the research sites and 1 km from the river's edge inland, consists of 45% lightly to moderately disturbed, tall forest; the remainder is mostly current or re-grown farm clearings. Lightly disturbed, tall forest prevails further from the villages, but within foraging range. The sample included all 16 households in Yapuwas and 16 (of about 53) of the households in Krausirpe.

Five students carried out field work from June 1994 to December 1996. During days chosen at random each month, students identified, measured, weighed and valued all goods entering households from 6 am to 6 pm, and noted the place of origin of each item. Out of the 140 days during which we measured consumption, 71 took place in Yapuwas and 69 in Krausirpe, 54 in 1995 and 86 in 1996. We valued goods in the state of processing the goods were in when they entered the household. We used the buying price paid by the household to value the good. For goods without a village price,

once a month we asked a group of Tawahka to reach consensus on how much of a good with a known price (for example, salt) they would exchange in the village for the forest good without a price. We interviewed subjects to estimate the source and level of cash income earned during the 30 days before the interview. During 1995 and 1996 we carried out nine and eleven income surveys, respectively. We converted domestic values to international values using the yearly exchange rate and the purchasing power parity (PPP) index⁴. We did not collect prices in cities, so we cannot tell whether village prices mirror urban prices. As prices in remote villages are often lower than city prices and exchange rates and PPP indexes are based on urban prices, we may be underestimating forest values.

In the estimations, we included fish, game, wild plants, persistent crops in second-growth forest, firewood and timber brought to the household or sold, but excluded domesticated animals, minerals, water, industrial goods and cultivated crops. The omission of water underestimates the value of the forest because water might be the most important good or service many forested areas provide⁴. As people sold forest goods (for example, canoes or firewood) to outsiders and each other, we also estimate and report the gross value of the forest ha⁻¹ yr⁻¹ from the sale of forest goods¹². We estimate separate values of the forest from the consumption and from the sale of goods because we could not link a good brought from the forest now with the later sale of the same good; we thus avoid double counting. Nevertheless, we add the two values to estimate a potential upper range and ensure robustness in empirical results.

We assumed a foraging radius of 3 km from the village on the basis of all-day focal follows of foragers, but subtracted 2.4 ha of farmlands per household to adjust for lands under cultivation^{11,13}. We also subtracted a total of 5 ha for Yapuwas and 19 ha for Krausirpe to adjust for the area of the nucleated settlement. We expanded consumption values by 365/x and cash income by 12/y, where x and y are the number of days or months in a year in which we measured consumption or income. We report gross values because we did not collect information on foraging or manufacturing costs.

Table 1 contains estimates of the gross nominal value of all forest goods brought during weigh days and of the sale of all forest goods during selected months. We present the information for each village by year. For example, during the 42 days of 1996 when we measured consumption, people in the village of Krausirpe brought 1,551 goods from the forest; the value of those goods was US\$22,246. During 11 months of 1996, people from Krausirpe sold forest goods worth US\$91,041. As the values in Table 1 refer only to some days or months, we standardized those values so that they cover an entire year, and we express the standardized values per hectare for each village. Drawing on the raw information from Table 1, we present estimates of the annual value of the forest per hectare per year for each of the two village, for 1995 and for 1996, in Table 2. Several findings from Table 2 merit discussion.

First, consumption values vary between villages and years. The value of the forest is 2–3 times lower in the poorer village of Yapuwas than in Krausirpe, in part because households in Krausirpe extracted 69% more goods. In 1995, the value of forest goods consumed by people in Yapuwas reached US\$2.5 ha⁻¹ but the value in Krausirpe reached US\$9.05 ha⁻¹. During 1995–1996, the nominal value of the forest measured through consumption remained constant at about US\$2.52 ha⁻¹ yr⁻¹ in Yapuwas, but fell by 33% in Krausirpe, from US\$9.05 to US\$6.04 ha⁻¹ yr⁻¹.

Second, the consumption values per hectare per year were relatively low: US\$2.52 in Yapuwas and US\$6.04–9.05 in Krausirpe. The average nominal annual value of consumption per hectare for both villages ranged from US\$4.29 (1996) to US\$5.78 (1995). As the PPP index for Honduras during 1995–1996 was about 3, the welfare value of the forest compared with that of other nations was three times higher than the estimates suggest.

Table 1 Gross, nominal dollar value of forest goods brought in or sold by Tawahka Amerindians, Honduras, 1995–1996

	obs	Consumption			obs	1996 val	wd	Sale		
		1995 val	wd					1995 mon	1996 val	mon
Yapuwas	640	4,834	27	927	9,832	44	4,219	8.5	3,920	11
Krausirpe	1,101	17,324	27	1,551	22,246	42	69,378	10	91,041	11
Both	1,741	22,150	54	2,478	32,079	86	73,597	9.25	94,961	11

obs, individual items brought during weigh days; val, gross value of consumption or sales in lempiras; wd, weigh days or number of days in which consumption measured; mon, number of months covered in surveys. International Monetary Fund average yearly exchange rate for 1995 (9.47) and 1996 (11.7) was used to convert lempiras to US dollars. Sample includes 16 households in each village.

Table 2 Gross, annual, nominal dollar value of a hectare of tropical rain forest among Tawahka Amerindians, Honduras, 1995–1996

Village	Cons.		1995 Sale		Both		Cons.		1996 Sale		Both	
	\$	P	\$	P	\$	P	\$	P	\$	P	\$	P
Yapuwas	2.5	7.9	0.2	0.7	2.7	8.6	2.5	7.4	0.1	0.4	2.6	7.8
Krausirpe	9.0	28.6	3.2	10.1	12.2	38.7	6.0	17.7	3.4	10.0	9.4	27.7
Both	5.7	18.2	1.7	5.4	7.5	23.7	4.2	12.5	1.7	5.2	6.0	17.7

Values from Table 1 were used to estimate values per hectare per year for each village as follows: for consumption, in \$ ha⁻¹ year⁻¹, (gross value of consumption using exchange rate or PPP × 365 wd⁻¹) / ((π(3 × 1,000)²) / (10,000 – (2.4 ha per household × no. of households in sample + village area in ha))); for sales, in \$ ha⁻¹ year⁻¹, (gross value of sales using exchange rate or PPP × 12 / (no. of mon)) / ((π × (3 × 1,000)²) / (10,000 – (2.4 ha per household × no. of households in sample + village area in ha))). Purchasing power parity (PPP) rates of 3.16 (1995) and 2.93 (1996) from World Bank. Cons., value for consumption; Sale, value from sales; Both, value of consumption and sale added; \$, value using exchange rate; P, value using PPP.

Third, the value of the forest measured by the sale of forest goods was low and much higher in Krausirpe than in Yapuwás, but did not change much between years. The value of the forest measured by the goods sold in either year averaged US\$0.19 ha⁻¹ yr⁻¹ in Yapuwás and US\$3.31 ha⁻¹ yr⁻¹ in Krausirpe. On average, the forest yielded US\$1.75 ha⁻¹ yr⁻¹ from the sale of forest goods. Expressed in terms of PPP, values were US\$5.21–5.45 ha⁻¹ yr⁻¹.

Last, if one adds the value of consumption to the value of sale of forest goods one obtains pooled estimates that range from US\$6.06 to US\$7.50 ha⁻¹ yr⁻¹ using the official exchange rate and from US\$17.79 to US\$23.72 ha⁻¹ yr⁻¹ using PPP conversion indexes. The estimates may be biased upwards if some of the goods sold were also valued earlier as consumption when they entered the household.

Our results suggest that in this part of Central America the annual value of a hectare of tropical rain forest measured by all the goods local people actually consume or sell from the rain forest lies at the lower end of previous estimates. In a previous review⁴ of studies valuing raw materials and food from the rain forest, average worldwide values (converted to 1994 dollars with an additional correction for purchasing power) for food production and for raw materials were found to be US\$32 (food) and US\$315 (raw materials) ha⁻¹ yr⁻¹. The ranges were US\$6–75 ha⁻¹ yr⁻¹ for food and US\$43–1,014 ha⁻¹ yr⁻¹ for raw materials, or US\$49–1,089 ha⁻¹ yr⁻¹ for both types of goods combined.

The value of old-growth or of fallow forest is lower than the estimates in Table 2 because the estimates in Table 2 lump goods from both types of forest, exclude foraging and manufacturing costs, and include goods from rivers, riverbanks, ponds and creeks. The value of non-timber forest goods is also lower than the values in Table 2 because the estimates in Table 2 include timber and logs.

At least three tentative conclusions can be drawn from our study. First, the low economic value of the rain forest to local people explains why they might clear forests for other uses even when they might have secure rights of property to land. Second, tropical rain forests are worth more for the global than for the local values they supply^{2–4}. The value of the forest in carbon sequestration, in storing biological diversity and in providing ecological services overshadows the value of the forest as a purveyor of physical goods¹⁰. The mean total value of rain forests (excluding food and raw materials) was previously suggested to be US\$1,660 ha⁻¹ yr⁻¹ (ref. 4). Compared with that estimate, the Tawahka receive a small share of global benefits. Last, although the rain forest provides a low

annual flow of value in consumption or in cash income to local people, it may still be important to them as a form of insurance, particularly when mishaps, such as crop losses, strike¹⁴. At those times, the forest may gain prominence as a safety cushion, allowing people to smooth consumption. But one must be cautious before attaching too much weight to the insurance value of the forest. Rural people can protect themselves against mishaps either by taking precautionary measures before shocks take place (for example, intercropping, plot scattering) or by relying on reciprocity, tolerated theft, or out-migration after shocks strike. The relative importance of the forest next to other forms of informal self-insurance still needs to be settled through empirical analysis. □

Received 15 November 1999; accepted 2 May 2000.

1. Panayotou, T. & Ashton, P. *Not by Timber Alone* (Island, Washington DC, 1993).
2. Lampietti, J. A. & Dixon, J. A. To see the forest for the trees: a guide to non-timber forest benefits. Environment Department Paper No. 013 (The World Bank, Washington DC, 1995).
3. Chomitz, K. M. & Kumari, K. The domestic benefits of tropical forests: A critical review. *World Bank Obs.* 13, 13–35 (1998).
4. Costanza, R. *et al.* The value of the world's ecosystem services and natural capital. *Nature* 387, 253–260 (1997).
5. Kumari, K. An environmental and economic assessment of forest management options: a case study in Malaysia. Environmental Economic Series Paper No. 026 (The World Bank, Washington DC, 1995).
6. Peters, C. M., Gentry, A. H. & Mendelsohn, R. O. Valuation of an Amazonian rainforest. *Nature* 339, 655–656 (1989).
7. Ruiz Pérez, M. & Arnold, J. E. M. (eds) *Current Issues in Non-Timber Forest Products Research* (Center for International Forestry Research, Bogor, Indonesia, 1996).
8. Godoy, R. A., Lubowski, R. & Markandya, A. A method for the economic valuation of non-timber tropical forest products. *Econ. Bot.* 47, 220–233 (1993).
9. Ros-Tonen, M. A. F. (ed.) *NTFP Research in the TROPENBOS Programme: Results and Perspectives*. (The Tropenbos Foundation Wageningen, The Netherlands, 1999).
10. Daily, G. C. (ed.) *Nature's Services. Societal Dependence on Natural Ecosystems*. (Island, Washington DC, 1997).
11. Godoy, R. *Indians, Rain Forests, and Markets: Theory, Methods, and Analysis* (Columbia Univ. Press, New York, in the press).
12. McSweeney, K. The canoe in the tree. *Equinox* 103, 78–86 (1999).
13. Godoy, R. *et al.* Household determinants of neotropical deforestation by Amerindians in Honduras. *World Dev.* 25, 977–987 (1997).
14. Godoy, R. & Wong, G. Does economic development increase vulnerability? A test of consumption smoothing among the Tawahka Indians of the Honduran rain forest. *J. Dev. Econ.* (submitted).

Acknowledgements

The programs of Cultural Anthropology and Human Dimensions of Global Change of the National Science Foundation (USA), the Conservation, Food & Health Foundation, and the Social Science Research Council (USA) financed this research. We thank K. Bawa, M. Boscolo, J. Dixon, J. Harvey and D. Pimentel for comments on the manuscript.

Correspondence and requests for materials should be addressed to R.G. (e-mail: leandra@sprintmail.com).

Parallel evolution of virulence in pathogenic *Escherichia coli*

Sean D. Reid, Corinne J. Herbelin, Alyssa C. Bumbaugh, Robert K. Selander & Thomas S. Whittam

Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802, USA

The mechanisms underlying the evolution and emergence of new bacterial pathogens are not well understood. To elucidate the evolution of pathogenic *Escherichia coli* strains, here we sequenced seven housekeeping genes to build a phylogenetic tree and trace the history of the acquisition of virulence genes. Compatibility analysis indicates that more than 70% of the informative sites agree with a single phylogeny, suggesting that recombination has not completely obscured the remnants of ancestral chromosomes¹⁻³. On the basis of the rate of synonymous substitution for *E. coli* and *Salmonella enterica* (4.7×10^{-9} per site per year³), the radiation of clones began about 9 million years ago and the highly virulent pathogen responsible for epidemics of food poisoning, *E. coli* O157:H7, separated from a common ancestor of *E. coli* K-12 as long as 4.5 million years ago. Phylogenetic analysis reveals that old lineages of *E. coli* have acquired the same virulence factors in parallel, including a pathogenicity island involved in intestinal adhesion, a plasmid-borne haemolysin, and phage-encoded Shiga toxins. Such parallel evolution indicates that natural selection has favoured an ordered acquisition of genes and the progressive build-up of molecular mechanisms that increase virulence.

To elucidate the evolution of virulence and pathogenic mechanisms, we studied 14 strains representing common clones of enteropathogenic *E. coli* (EPEC), an important cause of infantile diarrhoea in the developing world, and enterohaemorrhagic *E. coli* (EHEC), one of the primary food-borne pathogens in the industrialized world⁴. In addition, we examined strains of other Shiga-toxin-producing *E. coli* serotypes and laboratory strain K-12. Seven housekeeping genes spaced around the *E. coli* chromosome (Fig. 1) were chosen for nucleotide sequencing. Five of these loci (*icd*, *mdh*, *mtlD*, *pgi* and *aroE*) have been shown to be polymorphic by enzyme electrophoresis⁵. In addition to these genes, we sequenced *rpoS*, a gene encoding sigma factor 38, which is involved in stress response, and *arcA*, a gene encoding the regulatory protein involved in the control of aerobic respiration. In the 21 *E. coli* strains studied, variation per locus ranges between 1% and 8% at both the nucleotide and the amino-acid level (Fig. 1).

Our first step in the phylogenetic analysis was to use the multilocus sequence data to reconstruct the evolutionary history of the pathogenic clones. This is complicated by the fact that recombination, in addition to mutation, contributes to clonal divergence^{1,3}. In nature, as a bacterial lineage accumulates mutations, bits of the genome are replaced by recombination, resulting in DNA sequences that consist of a clonal frame (relic of the ancestral chromosome) interrupted by recombined segments². To detect the effects of recombination on sequence divergence, we analysed the compatibility matrix⁶ of polymorphic nucleotide sites in the housekeeping loci. Two sites are compatible if a phylogeny exists in which all nucleotide changes at the sites can be inferred to have occurred only once. Incompatible sites require several changes, indicating that these sites have experienced recombination or repeated mutation. Figure 2 plots the compatibility matrix between all pairs of 201 binary sites (informative sites that have only two different nucleotides⁶) identified in a total of 5,664 bases sequenced in 21 strains. The observed compatibility is 74%, compared with 37% for

randomized matrices in a star phylogeny, and the neighbourhood similarity (the fraction of adjacent squares of the same colour) is 77%. For comparison, this value significantly exceeds the neighbourhood score (67%) obtained for 1,000 randomized matrices where the positions of sites are shuffled each time.

The total compatibility can be separated into two components: a 'within-locus' component based on the comparison of pairs of sites at the same locus (the fraction of coloured squares within the triangular regions of Fig. 2), and a 'between-locus' component based on the comparison of sites at different loci (the fraction of compatible pairs in the rectangular regions in Fig. 2). In general, there is a positive relationship between the within- and between-locus compatibilities (Fig. 2b). Overall, 85% of the pairs of sites within a locus are compatible, compared with 71% of the pairs at different loci. This observation indicates that even sites that are widely spaced on the bacterial chromosome retain similar phylogenetic information.

The plot of the compatibility matrix reveals two regions with a high concentration of incompatible sites (Fig. 2); blocks such as these are the hallmarks of past recombination⁶. One region includes a 42-nucleotide stretch starting at codon 317 in *mtlD*. The removal of the ten polymorphic sites in this segment increases the average between-locus compatibility to 72% (Fig. 2b). The second region is a cluster of eight polymorphic sites at the 3' end of *icd*. These sites reside on a DNA segment directly downstream from the insertion site of a naturally occurring bacteriophage⁷. When this 129-base pair (bp) segment is excluded from the analysis, the within-locus compatibility for *icd* increases from 73% to 86% (Fig. 2b). The removal of both incompatible regions increases the overall compatibility of sites from 74% to 78%.

What is the quality of the phylogenetic signal? To address this question, we used the method of split decomposition, which does not force the data into a tree-like phylogeny. Instead, this method allows for and thus can detect conflicting phylogenetic information⁸. This method is particularly useful in situations where a bifurcating tree is an inappropriate evolutionary model, such as for viral quasispecies⁹ or bacterial populations with high rates of interstrain gene transfer¹⁰. In these cases, split decomposition results in a mesh-like network of interconnected nodes reflecting the reticulate nature of their evolution.

Split decomposition of the multilocus sequence data shows that the clonal frames of the pathogenic *E. coli* are connected in a network that is more tree-like than mesh-like (Fig. 3a). Six nodes are simple bifurcations and two nodes are multifurcations. There is

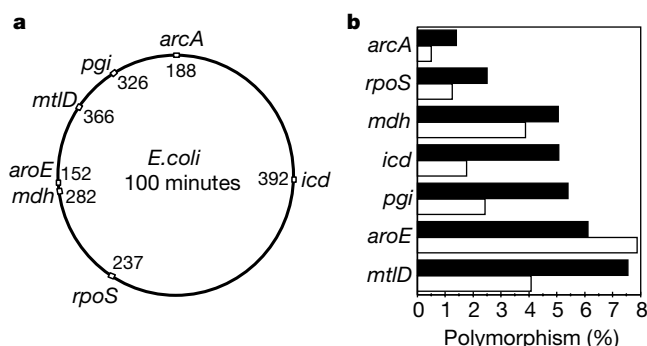


Figure 1 Genomic locations and variability of *E. coli* housekeeping genes. **a**, Location of 7 genes on the 100-minute map of the *E. coli* chromosome. The genes are clockwise as follows (protein, minute on map): *icd* (isocitrate dehydrogenase, 25), *rpoS* (sigma factor 38, 59), *mdh* (malate dehydrogenase, 73), *aroE* (shikimate dehydrogenase, 74), *mtlD* (mannitol 1-phosphate dehydrogenase, 91), *pgi* (phosphoglucose isomerase, 91) and *arcA* (aerobic respiratory control protein, 99). The number of codons sequenced is given next to each locus. **b**, The percentage of variable positions at the nucleotide (black bars) and amino-acid level (white bars).

only one case of parallel paths, indicative of a past recombinational exchange, which involves the relationship among E2348/69, DEC 2a, and the uropathogenic strain 536 (Fig. 3a).

Knowing that the underlying network is tree-like, we inferred a phylogeny for the strains that were based on sequences of the six loci (a combined total of 1,680 codons in 5 kilobases (kb)) that could be rooted with homologous genes from *Salmonella enterica*. A neighbour-joining tree revealed several main branches supported by bootstrap values greater than 85% (Fig. 3b). The same groupings, with trivial changes in topology, were found for the single most parsimonious tree, a unrooted phylogeny based on 7 loci which required a total of 383 mutations.

We compared the rooted phylogeny with the molecular clock hypothesis by the two-cluster method¹¹ to estimate the timescale involved. Nucleotide sequence data from 15 of the 21 strains fit the molecular clock; the most notable exception was strain K-12, which had a significantly longer branch than expected. To infer a timescale for the radiation of clones, we used the rate of synonymous substitution for *E. coli* and *S. enterica* of 4.7×10^{-9} per site per year³, which is based on a speciation date of 100 million years (Myr) ago³. This evolutionary rate for bacterial genes lies between the rates for nuclear genes of mammals (3.51×10^{-9}) and *Drosophila* (15.6×10^{-9})¹². The amount of divergence at synonymous sites indicates that the *E. coli* clonal frames trace back to a common ancestor that existed about 9.0 (8.1–9.9) Myr ago. It also indicates that the lineages destined to give rise to *E. coli* K-12 and O157:H7 separated about 4.5 (3.9–5.0) Myr ago.

The clonal phylogeny derived from multilocus sequencing is concordant with previous findings that were based on enzyme electrophoresis⁵. Three of the four pathogenic groups (EHEC 1, EHEC 2, and EPEC 1) are supported by bootstrap values greater than 85% (Fig. 2b). Members of the EPEC 2 group cluster together, but only two strains (DEC 11a and DEC 12a) have significant bootstrap support. The phylogenetic analysis also shows that the *E. coli* O55:H7 clone shares a recent common ancestor with *E. coli* O157:H7 (Fig. 3b) and is consistent with other features of the stepwise model for the evolution of this pathogen¹³.

Comparison of the distribution of specific genes that mark

Table 1 Virulence factors in pathogenic *E. coli* detected by PCR

Strain	Serotype	Chromosomal*			Plasmid†		
		<i>eae</i>	<i>stx1</i>	<i>stx2</i>	<i>ehly</i>	<i>katP</i>	<i>bfpA</i>
DEC5d	O55:H7	γ	-	-	-	-	-
5905	O55:H7	γ	-	+	-	-	-
493/89	O157:H-	γ	-	+	+	-	-
93-111	O157:H7	γ	+	+	+	+	-
OK-1	O157:H7	γ	+	+	+	+	-
921	O111:H9	+	-	-	-	-	-
2666-74	O26:H-	β	-	-	+	-	-
CL-37	O111:H8	+	+	-	-	-	-
DEC8b	O111:H8	+	+	+	+	-	-
90-1787	XO3:H-	-	-	+	-	-	-
CL-3	O113:H21	-	-	+	-	-	-
G5506	O104:H21	-	-	+	+	-	-
B2F1	O91:H21	-	-	+	-	-	-
B170	O111:H-	β	-	-	-	-	+
DEC12a	O111:H2	β	-	-	-	-	+
DEC11a	O128:H2	β	-	-	-	-	+
536	O6:H31	-	-	-	-	-	-
E2348/69	O127:H6	α	-	-	-	-	+
DEC 1a	O55:H6	α	-	-	-	+	+
DEC 2a	O55:H6	α	-	-	-	-	+

* Chromosomal genes include *eae* (intimin) on the LEE pathogenicity island, and *stx1* and *stx2* (Shiga-toxins 1 and 2) encoded by bacteriophages.

† Plasmid encoded genes include *ehly* (EHEC haemolysin) and *katP* (catalase) found on the EHEC plasmid (pO157) and *bfpA* (pilin subunit of bundle-forming pilus) encoded on the EAF plasmid.

‡ Carries a variant *ehly* allele (P. Feng, personal communication).

mobile elements associated with virulence (Table 1) supports the hypothesis that the high virulence of clones is a recent, derived state resulting from acquisition of virulence genes rather than an ancestral condition of primitive *E. coli*. The principal virulence factors and their mobile elements include the intimin gene (*eae*) encoded on a ~35-kb pathogenicity island called LEE (locus of enterocyte effacement); Shiga toxin genes (*stx1* and *stx2*) associated with bacteriophages; an enterohaemolysin gene (*ehly*) and catalase (*katP*), both of which occur on a ~90-kb EHEC plasmid¹⁴; and the main pilin subunit (*bfpA*) encoded on the EAF (EPEC adherence factor) plasmid. *Stx* genes and their bacteriophages are widely disseminated in the *E. coli* population¹⁵ and virtually identical *stx* sequences have been recovered from a variety of strains, indicating

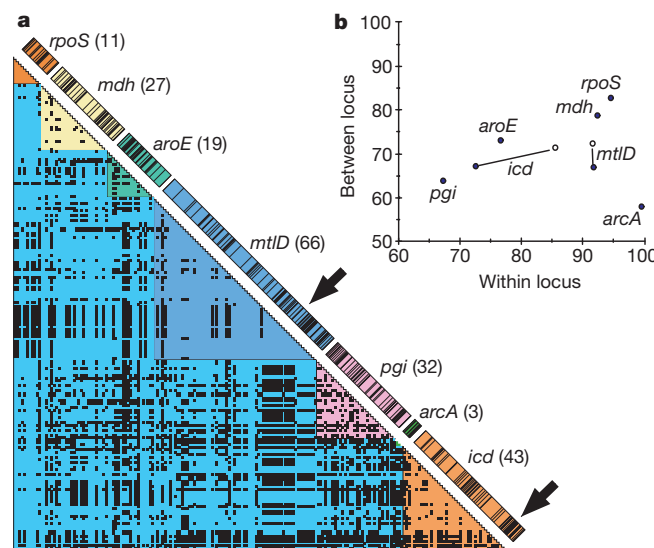


Figure 2 Compatibility of nucleotide polymorphisms. **a**, Lower triangle is a plot of all pairwise comparisons of 201 informative sites that are phylogenetically compatible within loci (solid colours) or between loci (blue). Incompatible pairs of sites are marked with a black square. The locations of the polymorphic sites within a locus are marked on the diagonal elements. For each locus, the number of informative sites is given in parentheses. The arrows mark two regions where incompatible sites within a locus are

clustered along the sequences. These regions include 10 informative sites in *mtlD* and 8 sites in the 3' end of *icd*. **b**, Average level of compatibility within and between loci shows that sites in *mdh* and *rpoS* have the greatest compatibility, whereas sites in *pgi* show the least compatibility. The lines show the effect of the removal of the incompatible regions in *icd* and *mtlD*.

recent, and presumably ongoing, toxin conversion in nature. The *ehlyA* sequences from *E. coli* O157:H7 (ref. 16) and a Shiga-toxin-producing O111 strain¹⁷ are also very similar, differing at 0.72% of the synonymous sites, even though these bacteria are on distant

branches (rate of synonymous substitution (d_s) = 6.2%) of the clonal phylogeny (Fig. 3b). For both phage-borne *Stx* genes and EHEC plasmid-borne *ehly* genes, the occurrence of nearly identical sequences in distantly related clones is indicative of recent horizontal transfer of these virulence elements.

There is no comparable evidence for recent mobility of intimin (*eae*) genes or the LEE pathogenicity island. Although the LEE is rare (<10%) in *E. coli* from natural hosts¹⁸, it appears to have been acquired several times. In pathogenic *E. coli* with α - or γ -intimin, LEE occupies a site in the *selC* gene¹⁹, whereas in strains with β -intimin, LEE is found in a different genomic site in *pheU*²⁰. It is noteworthy that α -, β - and γ -intimins are highly divergent in sequence (>15%), with mosaic allele structures suggesting that LEE has diversified *in situ* by mutation and recombination²¹.

The phylogenetic analysis suggests that gain and loss of mobile virulence elements has occurred several times, and in parallel in separate lineages of pathogenic *E. coli* (Fig. 3b). At least twice (Fig. 3b), this process involved the insertion of the LEE pathogenicity island into the chromosome, thereby setting the stage for acquisition of phage- and plasmid-borne genes. In two EHEC lineages, the process culminates with the addition of *Stx* genes and the EHEC plasmid encoding virulence factors implicated in haemorrhagic colitis. In contrast, EPEC lineages gained a different plasmid (EAF plasmid) specifying bundle-forming pili and other factors involved in infantile diarrhoea.

The parallel pattern of evolution suggests that there is a selective advantage favouring the build-up of specific combinations of virulence factors enabling the establishment and transmission of new virulent clones. Some pathogenic lineages, such as *E. coli* O157:H7, may be more likely than others to acquire foreign DNA because of an enhanced ability to recombine, a secondary effect of defective mismatch repair²². One intriguing possibility is that molecular interactions between mobile virulence elements lead to an ordered progression of change. This may be the case in cholera bacteria where a filamentous phage that forms pili critical for host colonization also acts as a receptor for a second phage carrying the cholera toxin gene²³. This type of direct interaction between mobile elements would promote the emergence of new pathogenic strains through the successive acquisition of virulence factors. □

Methods

Bacterial strains

Twenty pathogenic strains of *E. coli* were used as sources of DNA sequencing. The strains represent pathogenic groups identified previously by multilocus enzyme electrophoresis²⁴. The groups include enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC) and Shiga-toxin-producing *E. coli* (STEC). Serotypes are given in Table 1. Each strain is listed with country and year of original isolation given in parenthesis. EPEC 1: E2348/69 (UK 1969), DEC1a (USA 1956), DEC2a (Congo 1962). EHEC 1: 93-111 (USA 1993), OK-1 (Japan 1996), 493/89 (Germany 1989), 5906 (USA), DEC5d (Sri Lanka 1965). EHEC 2: DEC8b (USA 1986), CL-37 (Canada 1982), 2666-74 (USA 1974). EPEC 2: DEC11a (USA 1975), DEC12a (UK 1950), B170 (USA 1983); STEC: 90-1787 (Canada 1980s), CL-3 (Canada 1980), B2F1 (Canada 1985), G5506 (USA 1994). We also included strain 921 (Finland 1987), urinary tract isolate 536, and *E. coli* K-12 for purposes of comparison.

DNA sequencing

E. coli strains were grown on MacConkey's agar overnight at 37 °C. A single colony of each strain was transferred to 10 ml of Luria broth and grown overnight under agitation at 37 °C. Chromosomal DNA was isolated from each strain in accordance with the instructions in the Puregene DNA Isolation Kit from Gentra Systems (Minneapolis, Minn.). Isolated DNA was stored at 4 °C.

Oligonucleotide primers were designed on the basis of published sequences of house-keeping genes and virulence genes available in GenBank. Intimin alleles were detected by multiplex PCR²⁵. Primers were synthesized by a Beckman 1000 oligonucleotide synthesizer (Beckman, Fullerton, CA). Template DNA for cycle sequencing was obtained through amplification for 30 cycles as follows: 94 °C for 1 min, 53 °C for 2 min, and 72 °C for 2 min with an initial denaturing step of 94 °C for 5 min. The PCR products were purified with Qiaquick spin columns (QIAGEN, Valencia, CA) and suspended in the supplied elution buffer to suitable concentrations, as determined by agarose gel electrophoresis.

Cycle sequencing was performed with a Prism Ready Reaction DyeDeoxy Terminator Cycle Sequencing kit from Applied Biosystems. Sequencing gels were run on an Applied

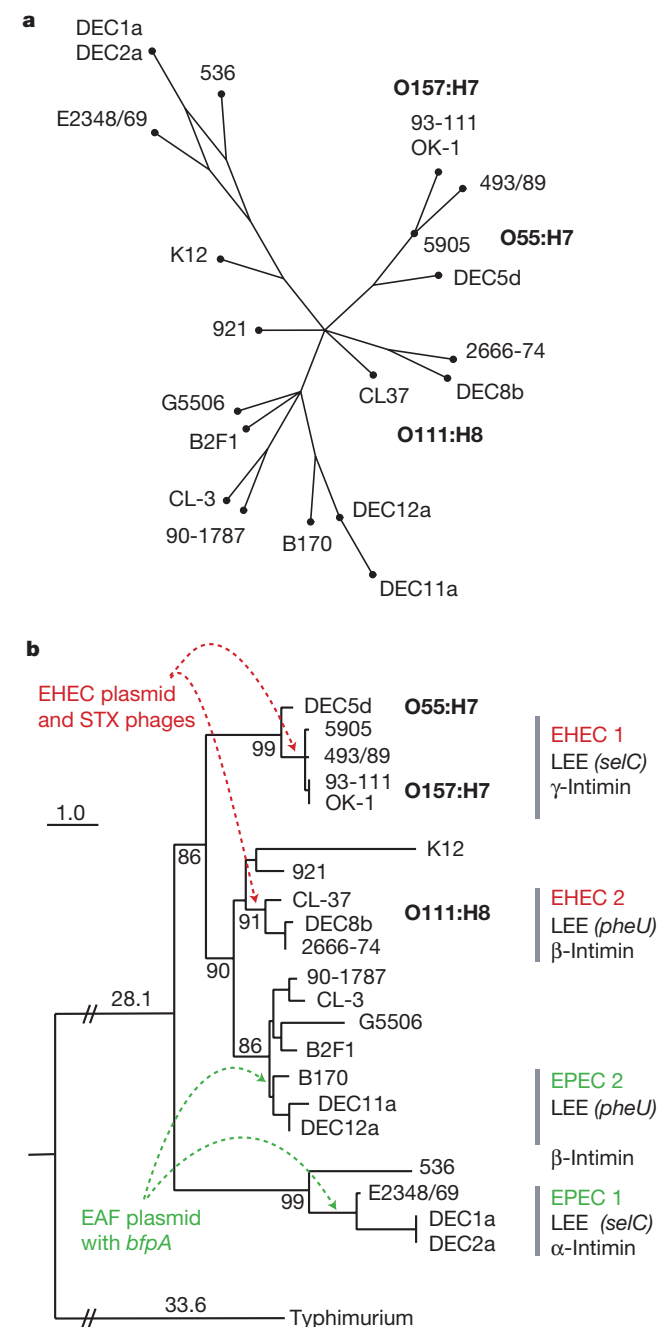


Figure 3 Phylogenetic analysis of 21 *E. coli* strains. **a**, Split decomposition graph of the relationship of strains based on concatenated alleles for 7 loci. The pairwise dissimilarity of the multilocus genotypes was estimated using hamming (uncorrected) distances. The graph was plotted with all edges set to equal length. **b**, Rooted phylogeny of pathogenic strains. The tree is based on concatenated sequences of 6 loci (a total of 5.0 kb) rooted with homologous sequences from *S. enterica* Typhimurium. The phylogeny was constructed with the neighbour-joining algorithm using the synonymous rate of substitution (d_s). The polymorphic codons implicated in recombination of *mtlD* for strain E2348/69 were excluded from the analysis. Branch lengths are measured in terms of the number of synonymous changes per 100 synonymous sites ($d_s \times 100$) and the numbers at the nodes are bootstrap confidence values based on 1,000 replicates. Arrows indicate the putative acquisition events for the mobile elements encoding the virulence factors listed in Table 1.

Biosystems 373A automated sequencer. Raw sequences of both DNA strands were analysed and concatenated by DNASTAR with additional internal sequencing primers designed based on the generated sequence data. All conflicting and putative polymorphic sites were sequenced at least three times to reduce sequencing error.

Phylogenetic analysis

Multiple-sequence alignment of the inferred amino-acid sequences was performed with Clustal W²⁶. A supergene was constructed for phylogenetic analysis by concatenating the individual gene sequences in the order: *rpoS*, *mdh*, *aroE*, *mtlD*, *pgi*, *arcA* and *icd*. The program Reticulate was used to identify putative regions of recombination or gene conversion through the construction of compatibility matrices⁶. A Monte Carlo approach was used to evaluate the significance of the matrices. Only binary parsimoniously informative sites were retained with incompatible regions removed from the analysis. Split decomposition was performed with the SplitsTree program²⁷.

Phylogenetic trees were inferred by the neighbour-joining algorithm using MEGA²⁸ or by the method of maximum parsimony using PHYLIP²⁹. The phylogeny based on 6 loci (combined total of 5,042 bp; *aroE* was excluded) was rooted with the homologous sequences from *S. enterica* extracted from the GenBank database. In the combined sequences, there were a total of 726 variable sites, 89 of which involved amino-acid replacements among the 22 sequences. Among the 21 *E. coli* sequences, there were 216 polymorphic sites, 40 of which predict amino-acid differences. Bootstrap confidence values were based on 1,000 simulated trees. Linearized trees and tests of the molecular clock hypothesis were based on the LinTree programs¹¹. The rate of synonymous substitution (d_s) and its standard error was calculated using MEGA²⁸. Confidence intervals for the divergence times were calculated from the standard errors of the genetic distances. The divergence times for nodes in the rooted phylogeny (Fig. 3b) were based on pair-wise comparison of 93-111 and E2348/69 ($d_s \times 100 = 8.46 \pm 0.81$), and 93-111 and CL 37 ($d_s \times 100 = 4.18 \pm 0.56$) for the synonymous sites in 1,888 codons combined from 7 genes.

Received 15 November 1999; accepted 19 April 2000.

- Guttman, D. S. & Dykhuizen, D. E. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* **266**, 1380–1383 (1994).
- Milkman, R. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* 2nd edn (eds Neidhardt, F. C. et al.) 2663–2684 (American Society for Microbiology, Washington DC, 1996).
- Lawrence, J. G. & Ochman, H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl Acad. Sci. USA* **95**, 9413–9417 (1998).
- Nataro, J. P. & Kaper, J. B. Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* **11**, 142–201 (1998).
- Whittam, T. S. et al. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect. Immun.* **61**, 1619–1629 (1993).
- Jakobsen, I. B. & Eastale, S. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. *CABIOS* **12**, 291–295 (1996).
- Wang, F. S., Whittam, T. S. & Selander, R. K. Evolutionary genetics of the isocitrate dehydrogenase gene (*icd*) in *Escherichia coli* and *Salmonella enterica*. *J. Bacteriol.* **179**, 6551–6559 (1997).
- Bandelt, H. J. & Dress, A. W. M. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Mol. Phylogenet. Evol.* **1**, 242–252 (1992).
- Dopazo, J., Dress, A. & von Haeseler, A. Split decomposition: a technique to analyze viral evolution. *Proc. Natl Acad. Sci. USA* **90**, 10320–10324 (1993).
- Holmes, E. C., Urwin, R. & Maiden, M. C. The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Mol. Biol. Evol.* **16**, 741–749 (1999).
- Takezaki, N., Rzhetsky, A. & Nei, M. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**, 823–833 (1995).
- Li, W.-H. *Molecular Evolution* (Sinauer Associates, Sunderland, Massachusetts, 1997).
- Feng, P., Lampel, K. A., Karch, H. & Whittam, T. S. Genetic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. *J. Infect. Dis.* **177**, 1750–1753 (1998).
- Karch, H. in *Escherichia coli* O157:H7 and Other Shiga Toxin-Producing *E. coli* Strains (eds Kaper, J. B. & O'Brien, A. D.) 183–194 (ASM, Washington DC, 1998).
- Newland, J. W. & Neill, R. J. DNA probes for Shiga-like toxins I and II and for toxin-converting bacteriophages. *J. Clin. Microbiol.* **26**, 1292–1297 (1988).
- Schmidt, H., Beutin, L. & Karch, H. Molecular analysis of the plasmid-encoded hemolysin of *Escherichia coli* O157:H7 strain EDL 933. *Infect. Immun.* **63**, 1055–1061 (1995).
- Schmidt, H. & Karch, H. Enterohemolytic phenotypes and genotypes of shiga toxin-producing *Escherichia coli* O111 strains from patients with diarrhea and hemolytic-uremic syndrome. *J. Clin. Microbiol.* **34**, 2364–2367 (1996).
- Boyd, E. F. & Hartl, D. L. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J. Bacteriol.* **180**, 1159–1165 (1998).
- Wieler, L. H., McDaniel, T. K., Whittam, T. S. & Kaper, J. B. Insertion site of the locus of enterocyte effacement in enteropathogenic and enterohemorrhagic *Escherichia coli* differs in relation to the clonal phylogeny of strains. *FEMS Microbiol. Lett.* **156**, 49–53 (1997).
- Sperandio, V. et al. Characterization of the locus of enterocyte effacement (LEE) in different enteropathogenic *Escherichia coli* (EPEC) and Shiga-toxin producing *Escherichia coli* (STEC) serotypes. *FEMS Microbiol. Lett.* **164**, 133–139 (1998).
- McGraw, E. A., Li, J., Selander, R. K. & Whittam, T. S. Molecular evolution and mosaic structure of α , β , and γ intimins of pathogenic *Escherichia coli*. *Mol. Biol. Evol.* **16**, 12–22 (1999).
- LeClerc, J. E., Li, B., Payne, W. L. & Cebula, T. A. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **274**, 1208–1211 (1996).
- Karaolis, D. K., Somara, S., Maneval, D. R., Jr., Johnson, J. A. & Kaper, J. B. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**, 375–379 (1999).
- Whittam, T. S. in *Escherichia coli* O157:H7 and other Shiga toxin-producing *E. coli* strains (eds Kaper, J. B. & O'Brien, A. D.) 195–209 (ASM, Washington DC, 1998).

- Reid, S. D., Betting, D. J. & Whittam, T. S. Molecular detection and identification of intimin alleles in pathogenic *Escherichia coli* by multiplex PCR. *J. Clin. Microbiol.* **37**, 2719–2722 (1999).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Huson, D. H. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73 (1998).
- Kumar, S., Tamura, K. & Nei, M. MEGA: Molecular evolutionary genetics analysis, version 1.0. (Pennsylvania State Univ., Univ. Park, Pennsylvania, 1993).
- Felsenstein, J. PHYLIP—Phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166 (1989).

Acknowledgements

The authors thank S. Plock for technical assistance. This research was supported by NIH grants (to T.S.W. and R.K.S.), and the Enteric Pathogen Research Unit at the University of Maryland Medical School.

Correspondence and requests for materials should be addressed to T.S.W. (e-mail: tsw1@psu.edu). Preliminary sequence data for *S. enterica* Typhimurium was obtained from The Institute for Genomic Research website at <http://www.tigr.org>.

Negative genetic correlation between male sexual attractiveness and survival

Robert Brooks

School of Tropical Biology, James Cook University, Townsville 4811, Queensland, Australia

Indirect selection of female mating preferences may result from a genetic association between male attractiveness and offspring fitness^{1,2}. The offspring of attractive males may have enhanced growth^{3–5}, fecundity^{3,4}, viability^{5–8} or attractiveness^{4,9–11}. However, the extent to which attractive males bear genes that reduce other fitness components has remained unexplored. Here I show that sexual attractiveness in male guppies (*Poecilia reticulata*) is heritable and genetically correlated with ornamentation. Like ornamentation^{12–14}, attractiveness may be substantially Y-linked. The benefit of mating with attractive males, and thus having attractive sons, is opposed by strong negative genetic correlation between attractiveness and both offspring survival and the number of sons maturing. Such correlations suggest either antagonistic pleiotropy between attractiveness and survival or linkage disequilibrium between attractive and deleterious alleles. The presence of many colour pattern genes on or near the non-recombining section of the Y chromosome may facilitate the accumulation of deleterious mutations by genetic hitchhiking^{15,16}. These findings show that genes enhancing sexual attractiveness may be associated with pleiotropic costs or heavy mutational loads.

The genetic relationships between male ornamentation, attractiveness to females and several fitness components in guppies were investigated. Female mate choice in this species is based on polymorphic male colour patterns¹⁷, and body³ and tail size. Male attractiveness was significantly heritable owing to additive genetic contributions from sires (44 sires, 98 dams, 280 offspring; $h_s^2 = 0.596 \pm 0.28$; $P = 0.044$) but not from dams (nested within sire) ($h_D^2 = 0.101 \pm 0.26$; $P = 0.71$; combined sire and dam h^2 estimate, 0.348 ± 0.15 , $CV_A = 29.75$; $h_s^2 > h_D^2$, $P = 0.05$, randomization test). Furthermore, there was a significant genetic correlation between attractiveness and ornamentation (Fig. 1), implying a common genetic basis, probably because attractiveness is based on ornamentation.

Predation is generally thought to be the most important determinant of juvenile guppy mortality in the field (12–24% survive to

maturity)¹⁸. However, under the rearing conditions imposed in these experiments, mean survival to maturity was 65.2% (s.e. = 24%, $n = 98$ litters), indicating that non-predatory juvenile mortality may be an important fitness component. Survival to maturity was negatively correlated with attractiveness (Table 1). This is due to genetic covariation among sires, rather than dam genetic, maternal or environmental effects (described by the positive, non-significant environmental correlation). Four lines of evidence indicate that the genetic component of this result is due to mortality of sons: (1) the genetic correlation between the number of sons maturing and their attractiveness was comparable with that between offspring survival and attractiveness; (2) there was a strong, negative correlation between male attractiveness and the survival of adult males; (3) the number of daughters maturing and the proportion of those surviving to the end of the experiment were not significantly genetically correlated with male attractiveness; and (4) male survival, both before and after maturity, is also negatively correlated with the ornamentation that females find attractive (Fig. 2).

Females that mate with an attractive male will benefit if his genes enhance the attractiveness of her sons^{4,9–11}, or the survival or fecundity of her offspring^{3–8}. These genetic benefits underpin fisherian and indicator models of indirect selection on mating preferences, respectively^{1,2}. A female guppy mating with an attractive and highly ornamented male will benefit indirectly by bearing attractive sons, but they will die earlier, at least under conditions similar to those imposed here. This result contradicts the predictions of indicator models and the findings of a recent meta analysis⁸, which indicate that more attractive or highly ornamented males sire offspring with higher survival probabilities. Only the most extreme handicap versions of indicator models could be consistent with this result, and only then if a positive genetic correlation between attractiveness and another fitness component negates the viability costs of attractiveness². The weak genetic correlations between male attractiveness and female survival and fecundity (Table 1) fail to provide such support. However, it is possible that under different environmental conditions such a positive genetic correlation may exist.

The results presented here are, however, consistent with Fisher's prediction that the fitness costs of ornamentation will evolve to balance the benefits of attractiveness¹. Such genetic costs are indicated by the findings that sons of successful male *Drosophila robusta*⁹ and *Gryllus bimaculatus*¹¹ not only have higher mating success themselves, but also develop more slowly, ostensibly impairing fitness. Additionally, artificial selection on comb size in roosters results in a correlated decrease in male viability¹⁹, suggesting a tradeoff between viability and this attractive ornament. Although these examples indicate a negative association between mating success and a fitness component, my data provide the first formal

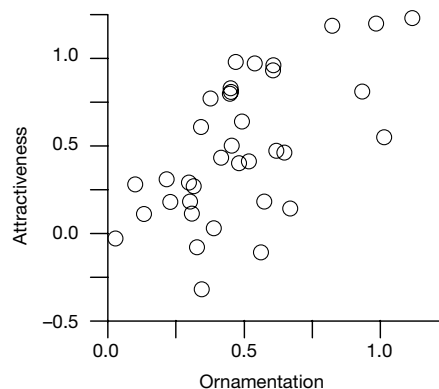


Figure 1 Approximate multiple genetic correlation²⁸ between ornamentation and male attractiveness. $R_G = 0.612 \pm 0.34$ ($P = 0.011$). Each point is the mean of sons of one sire. The model that minimized C_p incorporated tail area ($\beta = 0.42$, $P = 0.01$), brightness ($\beta = 0.35$, $P = 0.04$) and chroma ($\beta = 0.25$, $P = 0.11$) of orange spots, area of iridescence ($\beta = 0.23$, $P = 0.13$) and total number of spots ($\beta = -0.25$, $P = 0.14$). Body area and areas of orange, black and fuzzy black colour were excluded. The graph plots attractiveness against the value predicted by the coefficients estimated in the model. $n = 36$ half-sib families.

estimate of the negative genetic correlation. Moreover, the failure to find a positive genetic correlation between attractiveness and a fitness component shows that the offspring of attractive males are likely to be less fit (apart from their enhanced attractiveness) than those of less attractive males. More extensive study in a range of species and environmental conditions is required to elucidate the extent to which attractiveness is negatively correlated with fitness components because it influences not only evolution by sexual selection, but also the extent to which mating preferences evolve by indirect selection.

Genetic fitness costs associated with attractiveness may be either direct (pleiotropic) consequences of ornament expression, or due to linkage disequilibrium between attractiveness genes and deleterious alleles. Male guppies from attractive families experienced high mortality before the onset of maturity, showing that it was not the presence of the ornaments that increased mortality. However, survival early in life might be pleiotropically impaired for other reasons, such as the acquisition of resources for later investment in ornaments. Under normal levels of recombination, linkage disequilibrium between attractive and deleterious genes is unlikely, but on the Y chromosome physical linkage is possible owing to reduced recombination.

Irrespective of the relative importance of pleiotropy and linkage disequilibrium in determining the genetic correlations reported here, it is likely that much of the genetic variation underlying these traits is associated with the Y chromosome. Guppies have large, cytologically indistinguishable X and Y chromosomes, with suppressed recombination near a major sex-determining locus on the Y chromosome¹³. At least twenty ornament genes are so close to this locus that they have never been found on the X chromosome. Males inherit a combination of these as a Y-linked supergene¹². At least 17 other genes recombine between the X and Y chromosomes, at rates that increase with distance from the non-recombining section^{12,20}. Only five autosomal colour pattern genes have been identified¹². Many of the described sex-linked colour patterns were identifiable on males in this study, who bore a strong resemblance to their fathers and full and half brothers. The importance of Y-linked genes in determining male coloration is supported by the fact that sires contribute significantly more to the quantitative inheritance of ornamentation than do dams¹⁴ (R.B. and J. A. Endler, manuscript in preparation). The fact that h_s^2 of attractiveness exceeded h_d^2 almost sixfold, shows that much of the genetic variation in attractiveness is

Table 1 Genetic and environmental correlations between male attractiveness and components of fitness

	Sires	Dams	R_G	P	R_E
Offspring survival to maturity	39	92	-1.359 (0.70)	0.023	1.245 (1.54)
Sons at maturity	39	92	-1.180 (0.54)	0.021	1.642 (2.02)
Daughters at maturity	39	92	-0.232 (0.32)	0.47	0.050 (0.08)
Daughter's litter size	33	58	0.491 (0.52)	0.22	-4.509 (5.31)
	n		R	P	
Survival of mature sons*	39		-0.309	0.056	
Survival of mature daughters*	39		-0.051	0.759	
Proportion of daughters giving birth*	39		0.133	0.42	

* Spearman's nonparametric correlation among means for offspring of each sire. (s.e. in parentheses)

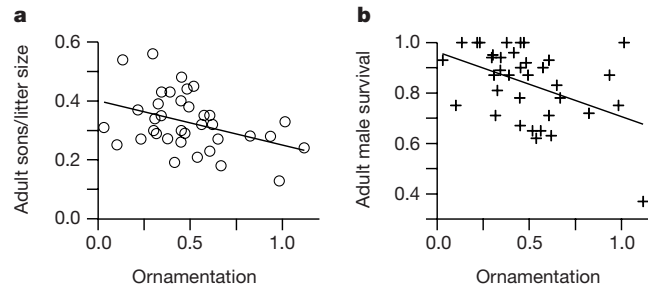


Figure 2 The relationship between male survival and attractive ornamentation (ornamentation in Fig. 1). **a**, Negative genetic relationship between the proportion of the original litter that survived to maturity and were male, and attractive ornamentation (product-moment correlation $R = -0.425$, $P = 0.010$). **b**, Genetic relationship

between attractive ornamentation and the survival of mature males to the end of the study (~6 months) (nonparametric correlation $R_s = -0.433$, $P = 0.008$). $n = 36$ sire means.

Y linked. This is to be expected, if the ornaments that females use to choose their mates are substantially Y linked.

The genetic correlations reported here are only due to genetic covariation among sires, but the extent to which the genes that reduce survival are Y linked cannot be estimated directly. However, the strong genetic correlations are due to physical linkage, pleiotropy or both. Thus, the genes that reduce survival are either close to or identical to those influencing ornamentation and attractiveness, and they will be Y linked to a similar extent. Antagonistic genes, which only benefit the heterogametic sex, but are otherwise costly, are more likely to accumulate on the Y chromosome^{15,21,22}.

Some deleterious alleles on the guppy Y chromosome are tightly linked with colour pattern genes²³. Their accumulation may explain why adult sex ratios in laboratory strains selected for Y-linked colour pattern markers become progressively female biased²⁴. Furthermore, an allele that is found on both X and Y chromosomes, but that does not frequently recombine between them, is associated with reduced vigour and viability only when inherited on the Y chromosome. In contrast, on the X chromosome²⁰ linked deleterious alleles are shed by recombination. The Y-linked sexually selected genes apparent in guppies may allow deleterious alleles to accumulate by genetic hitch-hiking¹⁶, resulting in the strong negative genetic correlations reported here. The possibility that sexual selection facilitates the hitch-hiking of deleterious alleles is of interest, because the accumulation of such genes is thought to drive the evolutionary reduction in size (and thus mutational target size) of Y chromosomes^{15,16,25}.

These data suggest that the only detectable benefit to females that

mate with attractive males is that they sire attractive sons, and that this benefit is opposed by pleiotropy or tightly linked genes that reduce offspring viability and therefore the number of adult sons. If such a balance between attractiveness and male mortality occurs in the wild, it may influence the strength of sexual selection, and the evolution of mating preferences. This dynamic may be more extreme in species with large Y chromosomes, such as guppies. □

Methods

Sires and dams were caught as juveniles from Alligator Creek (near Townsville, Australia) where they were introduced over 50 years ago. The half-sib design involved mating 44 sires to 3 virgin dams each (total 132 dams, one dam randomly allocated to each sire in each of 3 blocks 40 days apart). After 2 weeks, each dam was isolated in an 8-litre tank containing gravel and *Fontinalis* moss. After birth, offspring were counted and reared at initial densities of 7–11 fry per 8-litre tank. All reported measures were made on these F1 offspring. Half the water was replaced at intervals of about 30 days. The water was not aerated or filtered unless some of the fish looked unwell. Thus, both water quality and infection may have been responsible for fish mortality. As fish matured, they were separated by sex.

Measuring male attractiveness

Six males were assigned randomly (except that they were not full siblings) and individually to enclosures in a choice tank (Fig. 3). A previously mated control female was placed in each of the remaining two enclosures, and a virgin focal female was placed in the central area. This female was observed on the next morning for two 60-min bouts in a mirror suspended above the tank. Approach to within one body length of an enclosure was scored each minute: a male could score up to 120 on a given day (mean = 3.96, range = 0–45). Each set of six males was observed with 6–12 different females on consecutive days; males were reassigned to enclosures between days. Three hundred and nine focal females were used, none more than twice in total or once with any male. Attractiveness did not vary between tanks ($F_{[4,42]} = 1.86$, $P = 0.14$) but did between positions nested within the tank ($F_{[42,291]} = 4.31$, $P = 0.00$), probably owing to subtle lighting differences between enclosures. Residual attractiveness of each male's score was calculated from the mean score for the position. In the analyses presented, 'attractiveness' is the mean of these residual scores for a given male to the 6–12 females that saw him.

Estimating heritability of attractiveness

(Co)variances were estimated using restricted maximum likelihood (REML) (ASREML software²⁶). Owing to mortality, the statistical design is unbalanced. REML is more reliable than least squares estimation for unbalanced designs, and has the advantage of directly estimating the standard error of each (co)variance²⁷. Heritability of attractiveness was estimated in a two-factor model (sire, dam nested within sire), excluding block as a fixed factor because it influenced the likelihood of the model only negligibly. A randomization test was used to test whether $h^2_s > h^2_D$, keeping full sib families intact and the same number of dams per sire. Full sib families were randomized among sires and variance components calculated by least squares for 5,000 randomizations. Environmental or maternal effects, dominance or epistasis may inflate the estimated h^2_D of attractiveness, whereas the estimated h^2_s is free of these non-additive sources of variation²⁷. This makes my estimate of the extent to which h^2_s exceeds h^2_D conservative. However, intrabrood competition may increase the variance within full sib families, resulting in a lower estimate of h^2_D of attractiveness. The coefficient of variation in attractiveness among full brothers is uncorrelated with the number of potentially competing siblings ($r = 0.040$, $n = 80$ families, $P = 0.73$), or brothers, suggesting that such competition would not result in the observed sixfold reduction in the estimate of h^2_D .

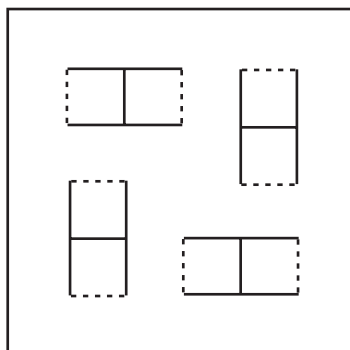


Figure 3 Scale plan of the choice tank used in measuring attractiveness, with eight enclosures. Outside walls covered in brown paper. Floor and side and median walls of enclosures (entire line) covered in light brown sand. Fronts of enclosures (dotted line) are transparent glass. Tank dimensions 42 × 42 cm.

Genetic correlations

There was only one measure of mortality per full-sib family, thus genetic (co)variances were estimated in a model with one random factor (sire) and block as a fixed factor. Correlations (top half of Table 1) were calculated using standard calculations for a half-sib design with one offspring per dam²⁷. R_G is due solely to genetic covariance among sires; the environmental correlation (R_E) includes genetic effects due to dam and dominance, epistasis and environmental sources of variation. Significance was tested using likelihood ratio tests²⁷. The correlations in the bottom half of Table 1 and in Figs 1 and 2 are among sire means calculated in two steps (mean of offspring of one dam, and then of all dams mated to the sire), with conventional significance tests. To calculate the multiple genetic correlation²⁸ between ornamentation and attractiveness (Fig. 1), multiple regression of (sire mean) attractiveness was fitted on ornamentation that minimized Mallow's C_p statistic among all possible models²⁹. R_G is the multiple correlation coefficient.

Daughter's fecundity

The proportion of daughters that gave birth from each full-sib family and the size of each litter were measured in a subsequent breeding experiment. Six males and six females were allowed to interact for 7 days, and then females were isolated until they gave birth or 4 months had elapsed.

Received 4 January; accepted 22 March 2000.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford Univ. Press, Oxford, 1930).
2. Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, 1994).
3. Reynolds, J. D. & Gross, M. R. Female mate preference enhances offspring growth and reproduction in a fish, *Poecilia reticulata*. *Proc. R. Soc. Lond. B* **250**, 57–62 (1992).
4. Moore, A. J. Genetic evidence for the "good genes" process of sexual selection. *Behav. Ecol. Sociobiol.* **35**, 235–241 (1994).
5. Petrie, M. Improved growth and survival of offspring of peacocks with more elaborate trains. *Nature* **371**, 598–599 (1994).
6. Norris, K. J. Heritable variation in a plumage indicator of viability in male great tits *Parus major*. *Nature* **362**, 537–539 (1993).
7. Sheldon, B. C., Merilä, J., Qvarnström, A., Gustafsson, L. & Ellegren, H. Paternal genetic contribution to offspring condition predicted by size of male secondary sexual character. *Proc. R. Soc. Lond. B* **264**, 297–302 (1997).
8. Møller, A. P. & Alatalo, R. V. Good-genes effects in sexual selection. *Proc. R. Soc. Lond. B* **266**, 85–91 (1999).
9. Etges, W. J. Sexual selection operating in a wild population of *Drosophila robusta*. *Evolution* **50**, 2095–2100 (1996).
10. Jones, T. M., Quinell, R. J. & Balmford, A. Fisherian flies: benefits of female choice in a lekking sandfly. *Proc. R. Soc. Lond. B* **265**, 1651–1657 (1998).
11. Wedell, N. & Tregenza, T. Successful fathers sire successful sons. *Evolution* **53**, 620–625 (1999).
12. Yamamoto, T. in *Handbook of Genetics* (ed. King, R. C.) (Plenum, New York, 1975).
13. Angus, R. A. in *Ecology and Evolution of Livebearing Fishes (Poeciliidae)* (eds Meffe, G. K. & Snelson, F. Jr) 51–68 (Prentice Hall, Englewood Cliffs, N.J., 1989).
14. Houde, A. E. Sex-linked heritability of a sexually selected character in a natural population of *Poecilia reticulata* (Pisces: Poeciliidae) (guppies). *Heredity* **69**, 229–235 (1992).
15. Rice, W. R. Evolution of the Y sex chromosome in animals. *BioScience* **46**, 331–343 (1996).
16. Rice, W. R. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**, 161–167 (1987).
17. Houde, A. E. *Sex, Color and Mate Choice in Guppies* (Princeton Univ. Press, Princeton, 1997).
18. Reznick, D. N., Butler, M. J., Rodd, F. H. & Ross, P. Life-history evolution in guppies (*Poecilia reticulata*). 6. Differential mortality as a mechanism for natural selection. *Evolution* **50**, 1651–1660 (1996).
19. von Schantz, T. et al. Artificial selection for increased comb size and its effects on other sexual characters and viability in *Gallus domesticus* (the domestic chicken). *Heredity* **75**, 518–529 (1995).
20. Nayudu, P. L. Genetic studies of melanin color patterns and atypical sex determination in the guppy *Poecilia reticulata*. *Copeia* **2**, 225–231 (1979).
21. Rice, W. R. Sexually antagonistic genes: experimental evidence. *Science* **256**, 1436–1439 (1992).
22. Roldan, E. R. S. & Gomendio, M. The Y chromosome as a battle ground for sexual selection. *Trends Ecol. Evol.* **14**, 58–62 (1999).
23. Haskins, C. P., Young, P., Hewitt, R. E. & Haskins, E. F. Stabilised heterozygosity of supergenes mediating certain Y-linked colour patterns in populations of *Lebistes reticulatus*. *Heredity* **25**, 575–588 (1970).
24. Farr, J. A. Biased sex ratios in laboratory strains of guppies, *Poecilia reticulata*. *Heredity* **47**, 237–248 (1981).
25. Rice, W. R. Degeneration of a nonrecombining chromosome. *Science* **263**, 230–232 (1994).
26. Gilmour, A. R., Cullis, B. R., Welham, S. J. & Thompson, R. ASREML software. (<http://res.bbsrc.ac.uk/aar/>) (1999).
27. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, Massachusetts, 1998).
28. Blows, M. W. & Allan, R. A. Levels of mate recognition within and between two *Drosophila* species and their hybrids. *Am. Nat.* **152**, 826–837 (1998).
29. Draper, N. & Smith, H. *Applied Regression Analysis* (Wiley, New York, 1981).

Acknowledgements

I thank E. Bolitho and Y. Williams for assistance, and M. W. Blows, M. J. Caley, L. Day, J. A. Endler, K. Hughes, M. D. Jennions, M. Petrie, J. D. Reynolds, F. H. Rodd and B. Walsh for discussion, encouragement, advice and comments on the manuscript. I am supported by a fellowship and grant from the ARC.

Correspondence and requests for materials should be addressed to R.B. (e-mail: rob.brooks@jcu.edu.au).

Identification of receptors for neuromedin U and its role in feeding

Andrew D. Howard^{*†}, Ruiping Wang^{‡†}, Sheng-Shung Pong^{*}, Theodore N. Mellin[§], Alison Strack[§], Xiao-Ming Guan^{*}, Zhizhen Zeng[‡], David L. Williams, Jr[‡], Scott D. Feighner^{*}, Christian N. Nunes[§], Beth Murphy[§], Judith N. Stair[§], Hong Yu^{*}, Qingping Jiang[‡], Michelle K. Clements[‡], Carina P. Tan^{*}, Karen K. McKee^{*}, Donna L. Hreniuk^{*}, Terrence P. McDonald[‡], Kevin R. Lynch^{||}, Jilly F. Evans[‡], Christopher P. Austin[‡], C. Thomas Caskey^{‡#}, Lex H. T. Van der Ploeg^{*} & Qingyun Liu[‡]

^{*} Department of Metabolic Disorders and [§] Department of Animal Pharmacology, Merck Research Laboratories, PO Box 2000, Rahway, New Jersey 07065, USA

[‡] Department of Pharmacology, ^{||} Merck Research Laboratories, West Point, Pennsylvania 19486, USA

[¶] Department of Pharmacology, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908, USA

[†] These authors contributed equally to this work

Neuromedin U (NMU) is a neuropeptide with potent activity on smooth muscle which was isolated first from porcine spinal cord and later from other species^{1–8}. It is widely distributed in the gut and central nervous system^{9,10}. Peripheral activities of NMU include stimulation of smooth muscle¹, increase of blood pressure¹, alteration of ion transport in the gut¹¹, control of local blood flow^{12,13} and regulation of adrenocortical function¹⁴. An NMU receptor has not been molecularly identified. Here we show that the previously described orphan G-protein-coupled receptor FM-3 (ref. 15) and a newly discovered one (FM-4) are cognate receptors for NMU. FM-3, designated NMU1R, is abundantly expressed in peripheral tissues whereas FM-4, designated NMU2R, is expressed in specific regions of the brain. NMU is expressed in the ventromedial hypothalamus in the rat brain, and its level is significantly reduced following fasting. Intracerebroventricular administration of NMU markedly suppresses food intake in rats. These findings provide a molecular basis for the biochemical activities of NMU and may indicate that NMU is involved in the central control of feeding.

We found a genomic sequence fragment with significant homology to FM-3 using the FAST_PAN program¹⁶ and a G-protein-coupled receptor (GPCR) query set. Cloning and subsequent analysis of this putative gene (referred to as FM-4) revealed an open reading frame of 1,236 base pairs (bp) encoding a protein of 412 amino acids with 51% identity to human FM-3 (Fig. 1a, b). We also isolated the rat orthologues of FM-3 and FM-4. The presumed open reading frame of rat FM-3 encodes a protein of 402 amino acids that is 80% and 73% identical to mouse and human FM-3, respectively (Fig. 1a, b). Rat FM-4 encodes a protein of 395 amino acids that is 75% identical to human FM-4 (Fig. 1a, b). All the receptors have seven predicted α -helical transmembrane domains containing highly conserved motifs, as observed in other members of the rhodopsin GPCR family (Fig. 1a).

To identify endogenous ligands of FM-3 and FM-4, we screened cells expressing human FM-3 against a panel of known peptides. In the initial screening, the peptide rat NMU-23 evoked a strong bioluminescent signal. Detailed analysis showed that human FM-3-expressing cells showed robust, dose-dependent responses to human, rat and porcine NMU peptides (Fig. 2a and Table 1). FM-4-transfected cells showed strong, dose-dependent responses to the NMU peptides with comparable affinity (Fig. 2b and Table 1). Fluorometric imaging plate reader (FLIPR) assays also showed that cells transfected with human FM-3 or FM-4 showed dose-depen-

Present address: Cogen Biotech Ventures, 711 Louisiana Street, Ste 3350, Houston, Texas 77002, USA.

dent, robust responses to the NMU peptides (Fig. 2c, d and Table 1). Cells transfected with vector alone failed to respond to NMU (concentrations tested up to 10 μ M; data not shown).

We performed radioligand binding assays using 125 I-labelled rat NMU-23. Saturation binding analysis showed that human FM-3-expressing cells displayed high-affinity, saturable and specific binding (dissociation constant $K_d = 0.3$ nM and binding capacity $B_{max} = 7.4$ pmol per mg protein; single class of binding sites; Fig. 2e). Competition binding analysis demonstrated high-affinity binding of NMU peptides to human FM-3 (Fig. 2f and Table 1). The results of functional and binding analyses indicate that NMU is an endogenous ligand of both FM-3 and FM-4, which are hereby designated NMU1R and NMU2R, respectively.

Human NMU1R is expressed in a variety of peripheral tissues with relatively high levels in the pancreas, testis, small intestine, adrenal cortex, liver and stomach (Fig. 3a), consistent with high levels of NMU detected in the gut. Expression of NMU1R could not be detected in human brain by northern analysis (data not shown). *In situ* hybridization analysis showed that NMU1R is specifically expressed in goblet cells in the ileum of the small intestine (Fig. 3b), indicating that NMU may regulate the secretory activity of goblet cells and may affect mucosal function. Expression of NMU2R was mostly restricted to specific regions of the brain. Multi-tissue northern analysis of human NMU2R detected relatively weak expression in the substantia nigra and thalamus of the brain, and in the spinal cord, thymus, thyroid and testis. However, expression could not be detected in the cerebral cortex of the brain or in the heart, liver, skeletal muscle, intestine, pancreas, placenta or kidney (data not shown). In the rat brain, NMU2R is expressed in paraventricular nucleus (PVN) of the hypothalamus, along the wall of the third ventricle in the hypothalamus and in the CA1 region of the hippocampus (Fig. 3c, d). In contrast, expression of NMU1R RNA could not be detected in the rat brain by *in situ* hybridization under similar conditions.

We examined the distribution of NMU messenger RNA in rat brain and compared its expression in fasted and fed rats. NMU mRNA is discretely located, with the most abundant signals in the ventromedial hypothalamic regions (lateral arcuate nucleus and median eminence; Fig. 4a) and in the caudal brainstem (nucleus of solitary tract, area postrema, dorsal motor nucleus of the vagus and inferior olive; data not shown). Co-localization studies indicated that NMU and pro-opiomelanocortin (POMC) RNAs are expressed close to each other in the arcuate nucleus and median eminence but are not expressed in the same neurons (Fig. 4b).

The expression of NMU in the ventromedial hypothalamus was reduced significantly ($-33 \pm 5\%$, $n = 3$ per group, $P < 0.05$) in rats fasted for 48 h (Fig. 4c) when compared with fed rats. This fasting-related decrease in NMU mRNA in the arcuate nucleus is similar to that observed for two anorectic peptide-encoding genes also located in the arcuate nucleus, POMC and cocaine-amphetamine-regulated transcript (CART)^{17,18}. We also investigated the expression of hypothalamic NMU mRNA in *ob/ob* mice¹⁹, a model of obesity. The pattern of NMU expression in the mouse hypothalamus differs

from that in rats. NMU mRNA in the mouse is detected in the suprachiasmatic nucleus, and appears to be decreased in the *ob/ob* mice (50.5 ± 3.1 versus 33.7 ± 4.2 nCi per g tissue for the lean and *ob/ob* mice, respectively; $P < 0.05$, $n = 3$). Weak NMU hybridization signals were also seen in the dorsomedial hypothalamic nucleus.



Figure 1 Sequence comparison of FM-3 and FM-4. **a**, Alignment of FM-3 and FM-4 protein sequences from mouse, rat and human. The seven transmembrane domains (TM 1–7) are underlined. Asterisk, N-linked glycosylation sites. For rat FM-3, no in-frame ATG initiation codon was found so a noncanonical leucine codon may serve as the initiator. **b**, Dendrogram representation of FM-3, FM-4 and the type 1 neurotensin receptor NTSR1. The GenBank accession numbers for the sequences are human FM-3/GPR66 (NM_006056), rat FM-3 (AF242873), mouse FM-3 (NM_010341), human FM-4 (AF242874), rat FM-4 (AF242875) and human NTSR1 (X70070).

Table 1 Functional and binding affinities of NMU to human NMU1R and NMU2R

Peptides	Human NMU1R			Human NMU2R	
	EC ₅₀ (Aeq)	EC ₅₀ (FLIPR)	IC ₅₀	EC ₅₀ (Aeq)	EC ₅₀ (FLIPR)
Human NMU-25	1.0	0.4	2.7	1.0	1.8
Rat NMU-23	0.75	0.6	8.7	2.9	1.7
Porcine NMU-8	0.11	0.8	1.7	0.3	1.3
Porcine NMU-25	1.36	n.d.	1.8	2.8	n.d.

Half-maximal response concentrations (EC₅₀, in nM) were determined by the aequorin assay (EC₅₀ (Aeq)) in HEK293/aeq17 cells or the FLIPR assay (EC₅₀ (FLIPR)) in COS-7 cells. Half-maximal inhibition concentrations (IC₅₀, in nM) were determined by radioligand binding assays. n.d., not determined.

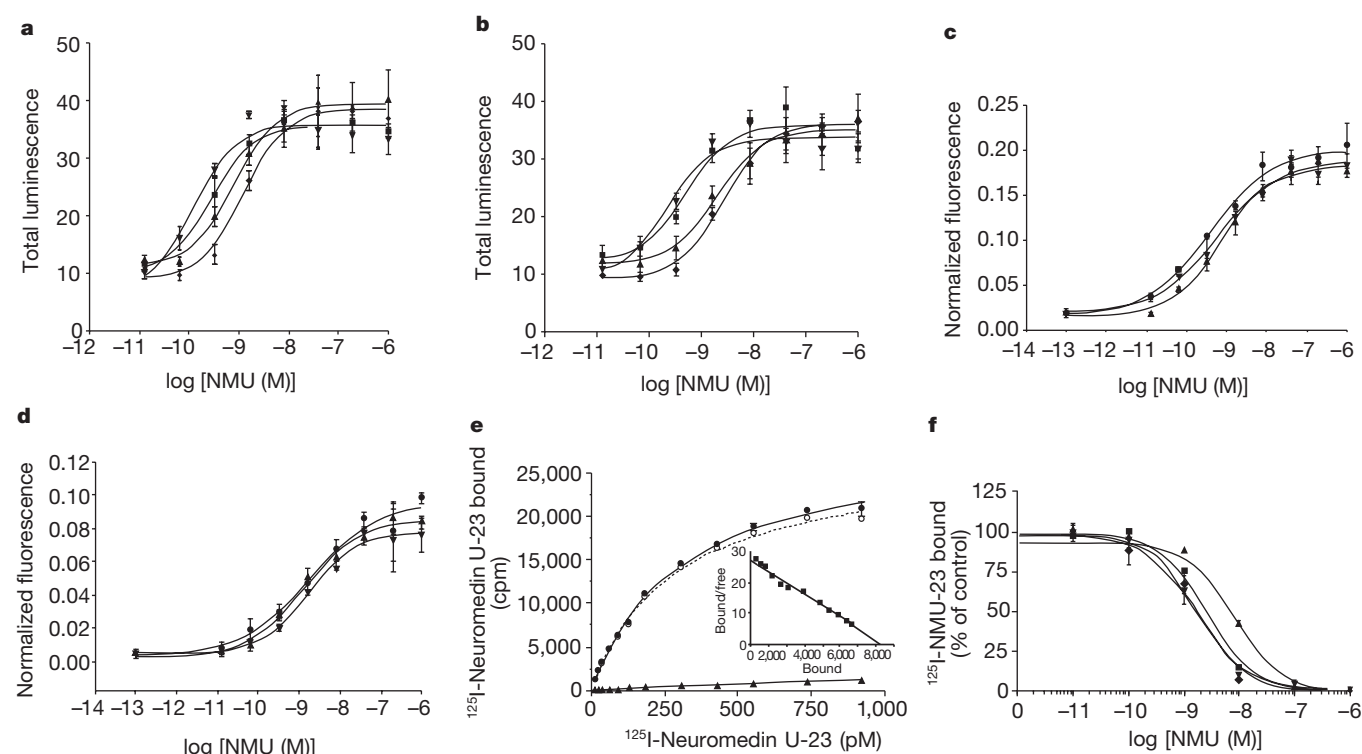


Figure 2 Functional and binding analysis. Activation by NMU in the aequorin assay of human FM-3 (**a**) and FM-4 (**b**), and in the FLIPR assay of human FM-3 (**c**) and FM-4 (**d**). Square, human NMU-25; up triangle, rat NMU-23; down triangle, porcine NMU-8; diamond, porcine NMU-23. **e**, Radioligand binding analysis with 125 I-labelled rat NMU by

saturation isotherm of human FM-3. Filled circle, total binding; triangle, nonspecific binding; open circle, specific binding. Inset, Scatchard analysis of the radioligand binding saturation isotherm. **f**, Competition binding analysis of human FM-3. Symbols as in **a–d**. Results shown are the means $\pm$ s.e.m. of triplicate determinations.

As NMU, like POMC and CART, is located in the arcuate nucleus in the rat brain, and all respond similarly to fasting, we tested whether NMU is also involved in the central control of feeding. Rats were injected with rat NMU-23 (1, 3 or 10 μ g, intracerebroventricular administration (ICV)) and monitored for overnight food intake and feeding duration. Injection of 3 or 10 μ g of NMU significantly decreased overnight food intake (analysis of variance

(ANOVA), $F(3)$ 8.4, $P = 0.0002$; Fig. 5a). Examination of feeding behaviour showed significant decreases in time spent feeding (cumulative feeding duration) by animals injected with NMU at both 3 μ g (67% of control) and 10 μ g (61% of control; Fig. 5b). Diet-induced obese (DIO) mice injected with 3 μ g of NMU also showed a significant reduction in overnight food intake ($-45 \pm 7\%$, $P < 0.01$).

In a second experiment, rats with radiotelemetric transmitters were monitored for core body temperature and gross motor activity, along with food intake, water intake and body weight after NMU administration (0, 1 or 3 μ g, ICV). As before, overnight food intake was decreased significantly. Additionally, water intake and body weight were suppressed in rats injected with 3 μ g NMU (data not shown). Core body temperature was transiently increased by NMU (3 μ g) during the first hour after NMU administration (Fig. 5c). The level of gross motor activity, compared with the pre-treatment level, was elevated ($P < 0.05$) during the first hour after injection with 1 or 3 μ g of NMU (Fig. 5d), but returned to normal for the remainder of the study. Such a transient increase in motor activity may account for only a small portion of the decrease in food intake as the suppression of feeding, in contrast to the altered motor activity, appears to be maintained for most of the study duration (Fig. 5b).

Nonspecific, toxic or aversive effects of NMU were ruled out by two tests, conditioned taste aversion and sodium appetite. In the conditioned taste aversion assay, saccharin was paired with NMU administration or with LiCl, a toxin that will cause rats to avoid saccharin in the future²⁰. NMU did not induce any decrease in saccharin intake, whereas LiCl caused severe taste aversion (Fig. 5e). In a second behavioural test, sodium appetite was measured. Rats have a robust appetite for salt and will drink avidly when sodium is depleted. However, when administered with LiCl or other toxins, rats will decrease sodium intake^{21,22}. In contrast to LiCl, NMU

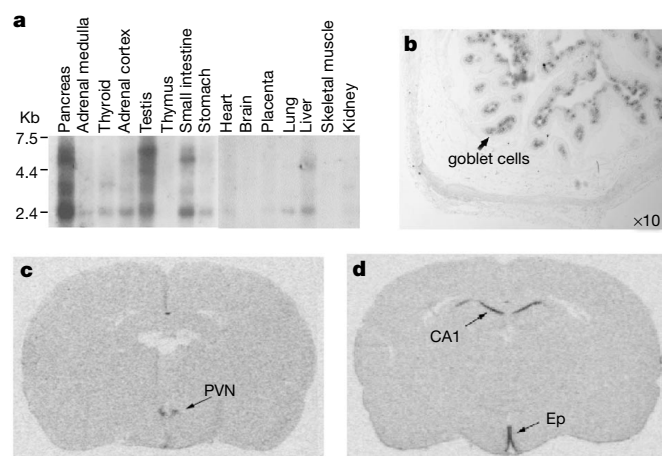


Figure 3 Expression of NMU1R and NMU2R. **a**, Northern analysis of human NMU1R. **b**, *In situ* hybridization analysis of NMU1R in monkey intestine. **c**, **d**, *In situ* hybridization analysis of NMU2R in the rat brain. PVN, paraventricular nucleus of the hypothalamus; Ep, ependymal layer in the wall of the third ventricle; CA1, CA1 layer of the hippocampus. The signals were completely blocked in the presence of 100-fold molar excess of unlabelled probe.

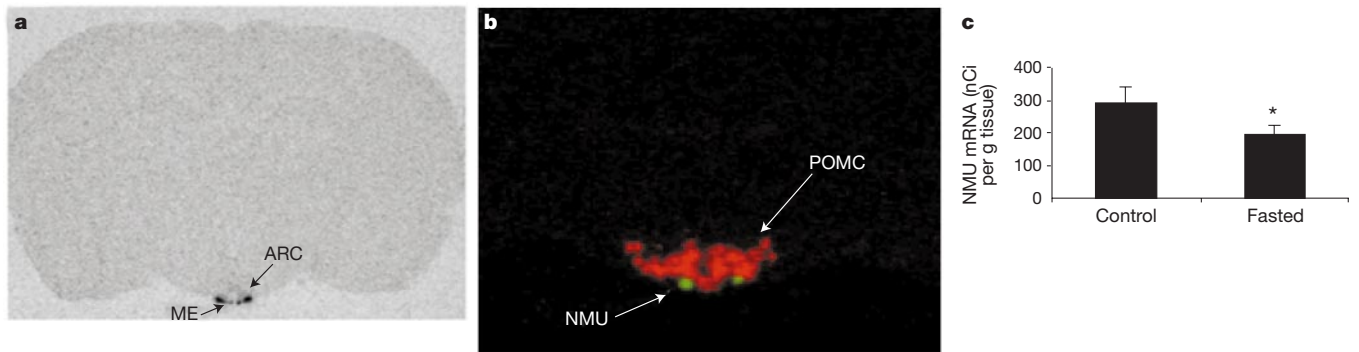


Figure 4 *In situ* hybridization analysis of NMU in the rat brain. **a**, Localization of NMU mRNA in coronal brain sections. ARC, arcuate nucleus; ME, median eminence. **b**, Localization of NMU and POMC mRNA in the rat hypothalamic arcuate nucleus and median eminence. Data represent two fused representative pseudocolour-coded (red,

POMC; green, NMU) images from a series of immediately adjacent 14- μ m coronal sections. **c**, Decrease in NMU mRNA in the ventromedial hypothalamic area in rats fasted for 48 h. Data shown are means $\pm$ s.e.m. of three experiments. Asterisk, $P < 0.05$; Student's *t*-test.

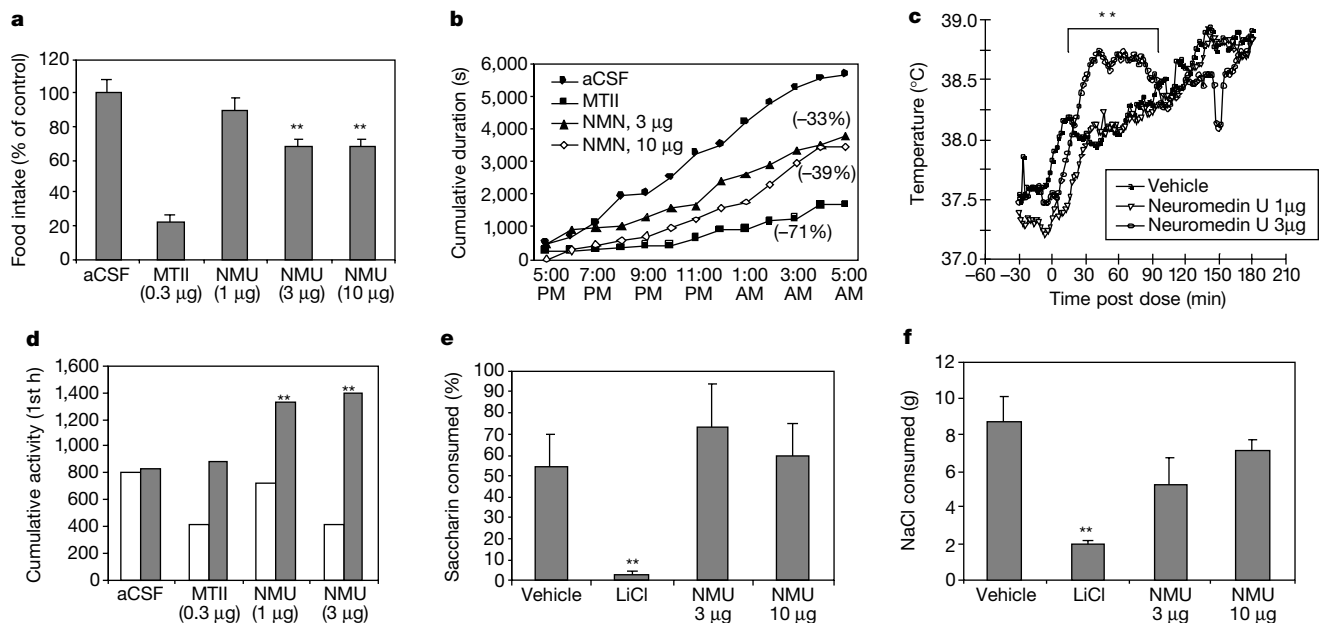


Figure 5 Effect of ICV-administrated NMU in rats. Reduction in food intake (**a**) and feeding duration (**b**) in rats injected with 3 μ g and 10 μ g of NMU, and with the melanocortin agonist MT-II (0.3 μ g; $t(28)10.2$, $P(0.01)$). Double asterisk, $P < 0.01$; $n = 11$ per group. **c**, Change in core body temperature. **d**, Change in gross motor activity in the first hour after

dosing. Double asterisk, $P < 0.02$, $n = 6$ per group. **e**, Conditioned taste aversion test. Double asterisk, $P = 0.02$; $n = 6$ per group. **f**, Sodium appetite test. Double asterisk, $P = 0.008$; $n = 6$ per group.

showed no significant effect on sodium appetite (Fig. 5f).

In conclusion, we have identified and characterized two receptors for NMU, bringing the total number of GPCRs that respond to the four classes (kassinin-, bombesin-, neurotensin-like and neuromedin U) of neuromedins to ten. We have also shown that NMU and NMU2R are located in hypothalamic areas consistent with the location of other peptides and receptors that control food intake. Our animal studies indicate that NMU may be a physiological regulator of food intake. The identification of two receptors for NMU and characterization of its central effect on feeding will stimulate further research into the biochemical and physiological functions of this brain-gut peptide. □

Methods

Molecular cloning

We cloned FM-4 by polymerase chain reaction (PCR) from human testis complementary DNA using two primers flanking the full-length coding sequence of FM-4. Partial

sequences of rat FM-3 and FM-4 were isolated using degenerate PCR from rat genomic DNA. We synthesized specific primers for each gene and used them to carry out genomic walking using the rat GenomeWalker kit (Clontech). Sequences corresponding to the start and stop codons were identified, and PCR was used to isolate the full-length coding sequences using primers flanking the open reading frames.

Functional and binding assays

We generated HEK 293/aeq17 (ref. 23) cells stably expressing human FM-3 as described²⁴ and performed aequorin bioluminescence assays as described^{23–25}. Eighty-three peptides (Phoenix Pharmaceuticals) were arrayed in a 96-well plate and used in the initial screening. The peptides included: ACTH, aldosterone secretion inhibiting factor (ASIF), α -MSH, amyloid β -protein (1–40), apelin-16, apelin-17, apelin-36, bombesin, insulin C peptide, cachectin 114–130, CART, cortistatin-17, cortistatin-29, DDAVP, delta sleep inducing peptide (DSIP), DTLET, endothelin-1, exendin-3, exendin-4, GRP, growth hormone releasing peptide-6, IRL-2500, kyotorphin, Leu⁸-phyllolitorin, MCH, morphiceptin, motilin, β -MSH, γ -MSH, neuropeptide NEL, neurokinin A, neurokinin B, neuromedin U, NPY, NPY (22–36), [Leu²¹, Pro²⁴]NPY (porcine), neurotensin, neuropeptide NGE, nocistatin, NPAF, NPFF, NPFF, orexin-A, orexin-B, proadrenomedullin-20 (PAMP), pancreastatin, pancreatic peptide PP, peptide YY, Phe⁸-phyllolitorin, POMC hinge peptide, prepro-7B2 (177–207), prepro-apelin (23–57), prepro-apelin (23–77),

prolactin-releasing peptide (PrP)-20, PrP-31, sarafotoxin S6a, sarafotoxin S6b, sarafotoxin S6c, secretoneurin, somatostatin 28 (1–12), substance P, thymosin β -10, enterostatin VPDPR, [β -Ala⁸]neurokinin A (4–10), [D-Tyr⁶, β -Ala¹¹, Phe¹³, Nle¹⁴]-Bombesin (6–14), [pGlu]apelin-13, enterostatin APGRP, agouti-related peptide (AGRP) 25–31, AGRP 54–82, AGRP 83–132, YMRP-amide and LPLRF-amide. Human NMU-25 was custom synthesized (Research Genetics). Porcine NMU-8, NMU-25 and rat NMU-23 were purchased from Phoenix Pharmaceuticals. FLIPR assays were performed in COS-7 cells transiently transfected with FM-3/pIRESpuromycin, FM-4/pcDNA3.1 or control vector as described²⁴. All transient transfections were carried out using Lipofectamine-2000 (GIBCO-BRL) by following the manufacturer's instructions. We carried out radioligand binding analysis as described²⁴. Rat NMU-23 was labelled with [¹²⁵I] at its amino-terminal tyrosine residue (Woods Assay) to a specific activity of ~ 2,000 Ci mmol⁻¹.

Expression analysis

Multi-tissue northern blots (Clontech) and *in situ* hybridization analysis of human NMU1R in monkey intestine was done as described²⁴. *In situ* hybridization analysis of NMU2R from rat brain was carried out using ³³P-labelled antisense oligonucleotide probes (equal mix of oligo 420, 5'-AGGAAAGGGTAATTGTGGCCACATCTCGTAGAT TTCCA-GAGGCATC-3', and oligo 421, 5'-CACAGTCTCGAAGAGGGCTGTCTTGAA GTAG-TATCCACACAGGC-3') as described²⁶. Quantitative *in situ* hybridization of NMU and POMC was performed using the same procedure with the NMU probe (5'-TTCTGGTGGTAATCTTTGAGGCGATATTGGCGTACCTCTGCAAGC-3') and the POMC probe (5'-CGTCTCTGATGATGGCGTCTTGAAGAGCGTCACAGGGGGCGTCT-3').

Animal studies

Male rats (Charles River Sprague Dawley) weighing 250–350 g were maintained in a temperature and humidity controlled facility with a 12-h light/dark cycle (04:00 lights on). Cannulation and ICV administration were performed essentially as described²⁷. Rats were injected ICV with 1, 3 or 10 μ g of rat NMU-23 (Phoenix Pharmaceuticals). Additional rats were injected ICV with either 0.3 or 0.03 μ g of MT-II (Peninsula Laboratories) as a positive control for food intake suppression. One group of rats also had a radio transmitter placed in the peritoneal cavity for measurement of core body temperature and gross motor activity (MiniMitter). In the mouse studies we used one-year-old, diet-induced obese (DIO, body fat content 60%) male C57BL/6 mice (Taconic). Mice were ICV-cannulated, injected and monitored for food intake as described above.

In the CTA study, rats were conditioned to 2 h daily access to water, with access to water from two bottles for 2 h each day for 3 days. On the fourth day, rats were given 0.15% saccharin for the 2-h period instead of water and saccharin consumption was measured. Rats were injected with NMU-23 (0, 3 or 10 μ g, ICV). LiCl was used as a positive control (0.15 M; 2 ml kg⁻¹, intraperitoneal (i.p.)). On the fifth day, rats were given saccharin alone for the first hour, then water was added for the remaining 23 h. We measured fluid consumption 1, 2 and 24 h after the injection. In the salt appetite assay, rats were given 0.5 M NaCl salt water to drink for 3 days along with food and regular water. After 3 days, two injections of furosemide (5 mg per 0.2 ml, subcutaneous) were given 1 h apart to sodium-deplete the rats. Rats were then returned to salt-free water and given a sodium-deficient diet. Twenty-four hours after furosemide administration, rats were given NMU (0, 3 or 10 μ g, ICV), or LiCl (0.15 M, 2 ml kg⁻¹, i.p.), and given water and 0.5 M NaCl to drink. Fluid consumption was measured 1, 2 and 24 h after dosing.

All rodent studies were conducted in accord with rules and guidelines of the Merck Research Laboratories Institutional Animal Care and Use Committee and the NIH Guidelines for the Care and Use of Laboratory Animals.

Received 16 March; accepted 17 May 2000.

- Minamino, N., Kangawa, K. & Matsuo, H. Neuromedin U-8 and U-25: novel uterus stimulating and hypertensive peptides identified in porcine spinal cord. *Biochem. Biophys. Res. Commun.* **130**, 1078–1085 (1985).
- Domin, J., Ghatei, M. A., Chohan, P. & Bloom, S. R. Characterization of neuromedin U like immunoreactivity in rat, porcine, guinea-pig and human tissue extracts using a specific radio-immunoassay. *Biochem. Biophys. Res. Commun.* **140**, 1127–1134 (1986).
- Conlon, J. M. *et al.* Primary structure of neuromedin U from the rat. *J. Neurochem.* **51**, 988–991 (1988).
- Minamino, N., Kangawa, K., Honzawa, M. & Matsuo, H. Isolation and structural determination of rat neuromedin U. *Biochem. Biophys. Res. Commun.* **156**, 355–360 (1988).
- Domin, J. *et al.* The distribution, purification, and pharmacological action of an amphibian neuromedin U. *J. Biol. Chem.* **264**, 20881–20885 (1989).
- O'Harte, F. *et al.* Primary structure and pharmacological activity of a nonapeptide related to neuromedin U isolated from chicken intestine. *Peptides* **12**, 809–812 (1991).
- Kage, R., O'Harte, F., Thim, L. & Conlon, J. M. Rabbit neuromedin U-25: lack of conservation of a posttranslational processing site. *Regul. Pept.* **33**, 191–198 (1991).
- Domin, J., Benito-Orfila, M. A., Nandha, K. A., Aitken, A. & Bloom, S. R. The purification and sequence analysis of an avian neuromedin U. *Regul. Pept.* **41**, 1–8 (1992).
- Honzawa, M., Sudoh, T., Minamino, N., Tohyama, M. & Matsuo, H. Topographic localization of neuromedin U-like structures in the rat brain: an immunohistochemical study. *Neuroscience* **23**, 1103–1122 (1987).
- Ballesta, J. *et al.* Occurrence and developmental pattern of neuromedin U-immunoreactive nerves in the gastrointestinal tract and brain of the rat. *Neuroscience* **25**, 797–816 (1988).
- Brown, D. R. & Quito, F. L. Neuromedin U octapeptide alters ion transport in porcine jejunum. *Eur. J. Pharmacol.* **155**, 159–162 (1988).
- Sumi, S. *et al.* Effect of synthetic neuromedin U-8 and U-25, novel peptides identified in porcine

- spinal cord, on splanchnic circulation in dogs. *Life Sci.* **41**, 1585–1590 (1987).
- Gardiner, S. M., Compton, A. M., Bennett, T., Domin, J. & Bloom, S. R. Regional hemodynamic effects of neuromedin U in conscious rats. *Am. J. Physiol.* **258**, R32–R38 (1990).
- Malendowicz, L. K. *et al.* Effects of neuromedin U (NMU)-8 on the rat hypothalamo-pituitary-adrenal axis. Evidence of a direct effect of NMU-8 on the adrenal gland. *Neuropeptides* **26**, 47–53 (1994).
- Tan, C. P. *et al.* Cloning and characterization of a human and murine T-cell orphan G-protein-coupled receptor similar to the growth hormone secretagogue and neurotensin receptors. *Genomics* **52**, 223–229 (1998).
- Retief, J. D., Lynch, K. R. & Pearson, W. R. Panning for genes—A visual strategy for identifying novel gene orthologs and paralogs. *Genome Res.* **9**, 373–382 (1999).
- Brady, L. S., Smith, M. A., Gold, P. W. & Herkenham, M. Altered expression of hypothalamic neuropeptide mRNAs in food-restricted and food-deprived rats. *Neuroendocrinology* **52**, 441–447 (1990).
- Kristensen, P. *et al.* Hypothalamic CART is a new anorectic peptide regulated by leptin. *Nature* **393**, 72–76 (1998).
- Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
- Chavez, M., Seeley, R. J. & Woods, S. C. A comparison between effects of intraventricular insulin and intraperitoneal lithium chloride on three measures sensitive to emetic agents. *Behav. Neurosci.* **109**, 547–550 (1995).
- Stricker, E. M. & Verbalis, J. G. in *Handbook of Behavioural Neurobiology of Food and Fluid Intake* (ed. Stricker, E. M.) 45–60 (Plenum, New York, 1990).
- Kissileff, H. R. Food-associated drinking in the rat. *J. Comp. Physiol. Psychol.* **67**, 284–300 (1969).
- Button, D. & Brownstein, M. Aequorin-expressing mammalian cell lines used to report Ca²⁺ mobilization. *Cell Calcium* **14**, 663–671 (1993).
- Liu, Q. *et al.* Identification of urotensin II as the endogenous ligand for the orphan G-protein-coupled receptor GPR14. *Biochem. Biophys. Res. Commun.* **266**, 174–178 (1999).
- Unguin, M. D., Singh, L. M., Stocco, R., Sas, D. E. & Abramovitz, M. An automated aequorin luminescence-based functional calcium assay for G-protein-coupled receptors. *Anal. Biochem.* **272**, 34–42 (1999).
- Guan, X. M., Yu, H. & Van der Ploeg, L. H. Evidence of altered hypothalamic pro-opiomelanocortin/neuropeptide Y mRNA expression in tubby mice. *Brain Res. Mol. Brain Res.* **59**, 273–279 (1998).
- Murphy, B. *et al.* Melanocortin mediated inhibition of feeding behavior in rats. *Neuropeptides* **32**, 491–497 (1998).

Acknowledgements

We thank M. Abramowitz and R. Stocco of Merck Frosst Canada for setting up the aequorin assay.

Correspondence and requests for materials should be addressed to Q.L. (e-mail: Jim_liu@merck.com). The GenBank accession numbers are human FM-3/GPR66 (NM_006056), rat FM-3 (AF242873), human FM-4 (AF242874) and rat FM-4 (AF242875).

Initiation of neural induction by FGF signalling before gastrulation

Andrea Streit*, Alyson J. Berliner†, Costis Papanayotou*,
Andrés Sirulnik* & Claudio D. Stern*†

* Department of Genetics and Development and † Center for Neurobiology and Behavior, Columbia University, 701 West 168th Street #1602, New York, New York 10032, USA

During neural induction, the 'organizer' of the vertebrate embryo instructs neighbouring ectodermal cells to become nervous system rather than epidermis. This process is generally thought to occur around the mid-gastrula stage of embryogenesis¹. Here we report the isolation of *ERNI*, an early response gene to signals from the organizer (Hensen's node). Using *ERNI* as a marker, we present evidence that neural induction begins before gastrulation—much earlier in development than previously thought. We show that the organizer and some of its precursor cells produce a fibroblast growth factor signal, which can initiate, and is required for, neural induction.

The prevailing model for neural induction suggests that cells differentiate into neural fates by default, but are normally inhibited by bone morphogenetic proteins (BMPs). The organizer, by emitting BMP antagonists, allows cells in its vicinity to execute their default neural programme^{2,3}. However, other work suggests a more

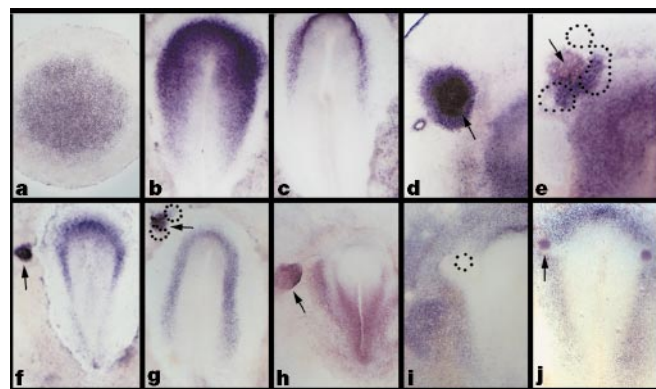


Figure 2 *cERNI* expression in the neural plate and its induction by Hensen's node and FGF8. *cERNI* is expressed in the prospective neural plate^{12,13} of embryos at the pre-primitive streak (a) and primitive streak (b) stage. It becomes restricted to the border at head-fold stages (c). Both Hensen's node (d, e; brown, arrow) and FGF8 (f, g; arrow) strongly induce *cERNI* expression, even in the presence of excess BMP4 (secreting cells outlined). FGF8 differentially regulates the expression of some BMP targets: it induces expression of *msx-1* within 3 h (h), but represses expression of the non-neural markers *GATA-2* (i) and -3 (j). Arrows in h and j and outline in i mark FGF8 beads.

portion of the FGF receptor¹⁶. SU5402 greatly reduced the frequency of induction by a grafted node of *Sox3* (from 16/16 in controls to 1/15—a 93% reduction; 5 h incubation; Fig. 3e) and of *cERNI* (from 20/20 in controls to 20/30—a 33% reduction in frequency; in the remaining embryos the intensity of expression was reduced; 3 h incubation; Fig. 3b). Moreover, cells secreting chimaeric FGF receptor markedly reduced induction by the node of *Sox3* (0/8; Fig. 3f) and of *cERNI* (6/38–83% reduction; Fig. 3c), whereas in the presence of control cells both genes were induced in 100% of cases ($n = 16$; Fig. 3a, d). We also noticed a marked reduction in the level of expression of both *cERNI* and *Sox3* in the normal neural domain of the host embryo in the presence of FGF inhibitors (compare Fig. 3k and j).

To investigate whether later steps of neural induction also depend on FGF signalling, we grafted quail nodes together with SU5402 beads into the area opaca of chick hosts. After overnight incubation (16 h), *Sox2* induction by the node was reduced to 20% (4/20; Fig. 3g, h) when compared with controls (10/10; Fig. 3g, i). The nodes, however, elongated normally and continued to express the organizer marker *chordin* (10/11; Fig. 3g–i), showing that the graft itself remains unaffected by the FGF inhibitor. These experiments strongly implicate FGF signalling as an essential component of the neural induction pathway initiated by the organizer.

In other systems, FGF and BMP signalling pathways can synergize or antagonize each other. We therefore investigated the effect of FGF8 on the expression of the BMP targets *msx1*, *GATA2* and *GATA3*. FGF8 induces *msx1* expression in 3 h (5/5; Fig. 2h), but represses expression of the non-neural markers *GATA2* (4/5; Fig. 2i) and -3 (4/6; Fig. 2j) in 5–6 h. The possibility remains that the ability of the node to induce *cERNI* and *Sox3* still requires BMP inhibition. To test this, we grafted stage 3⁺/4 nodes, or FGF8 beads, into host embryos together with cells secreting BMP4. Even in the presence of excess BMP4, both the node (Fig. 2e) and FGF8 (Fig. 2f) continued to induce *cERNI* (13/14 nodes; 10/12 FGF8) and *Sox3* (12/14 nodes; 6/8 FGF8). Together, these results suggest that the activity of FGF8 is complex, and cannot be explained merely through antagonism of the BMP pathway.

Both *cERNI* (Fig. 2a) and *Sox3* start to be expressed long before gastrulation, raising the possibility that the initial steps of neural induction could begin this early. At this stage, the organizer has not yet formed, but two cell populations that will contribute to it have been defined (Fig. 4a). The first is a group of middle layer cells

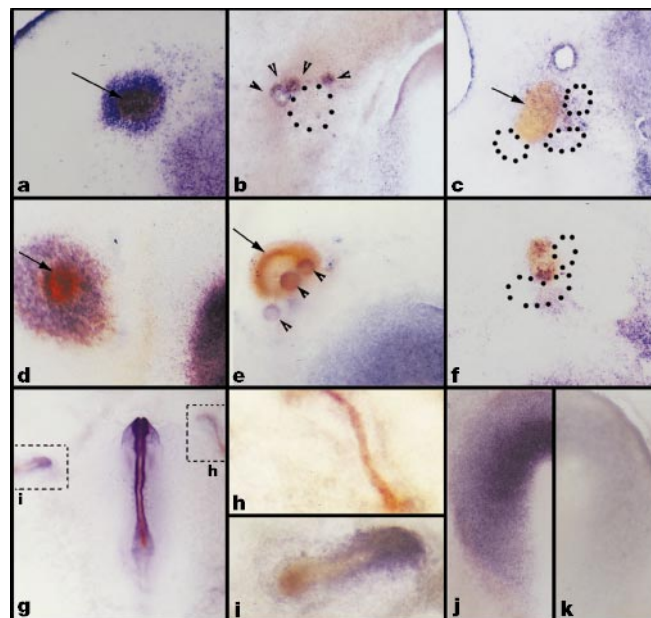


Figure 3 FGF signalling is an essential component of the neural induction pathway. Hensen's node (brown) induces *cERNI* (a, arrow), *Sox3* (d, arrow) and *Sox2* (left in g, i). The FGF receptor inhibitor SU5402 (arrowheads in b, e) inhibits induction of all three genes (b, e; right in g, h) by the node, which still elongates and expresses the organizer marker *chordin* (g–i; *Sox2* in purple, *chordin* in red). Cells secreting a soluble form of the FGF receptor (outlined) also greatly reduce induction of *cERNI* (c) and *Sox3* (f) by the node. The endogenous expression of *Sox3* is reduced in embryos treated with SU5402 (k) as compared with untreated embryos (j) (embryos processed simultaneously in the same vial).

associated with Koller's sickle (hereafter referred to as 'posterior cells'), which later moves to the centre of the embryo with the tip of the elongating primitive streak (stages 2–3)¹⁷. The second population, hereafter referred to as 'central cells', is located in the epiblast overlying the posterior cells until stage X, but moves to the centre of the area pellucida before primitive streak formation begins (stages XI–XIII)¹⁰. Hensen's node is formed when the two populations come together. We investigated the inducing ability of these two populations of organizer precursors, alone or in combination. Chick central cells grafted into the area opaca of a stage 3⁺/4 quail host do not induce *qERNI* (0/4), *Sox3* (0/17), *Sox2* (0/6) or the organizer marker *chordin* (0/5). In contrast, chick posterior cells induce *qERNI* (21/21; Fig. 4d) and *Sox3* (8/11; Fig. 4e) strongly after 5 h, but not the later neural marker *Sox2* or *chordin* (both after 16 h; $n = 10$). A combination of central and posterior cells induces both *Sox2* (3/5; Fig. 4f) and *chordin* (3/5; Fig. 4g) after overnight culture; other regions of area pellucida epiblast cannot substitute for the central cells in this assay ($n = 7$). These findings show that posterior, but not central cells, induce *ERNI* and *Sox3*. A combination of the two generates a functional organizer, as well as expression of the later neural marker *Sox2*.

Where is FGF8 expressed at these early stages? A stage X, the posterior cells, but not central cells, express FGF8 (Fig. 4b). As the hypoblast starts to form from the posterior margin, it also expresses FGF8, albeit at low level (Fig. 4c). By stage XIII, the domain of *cERNI* and *Sox3* expression in the epiblast mirrors the area covered by hypoblast. Consistent with these expression patterns, grafts of hypoblast into the area opaca of stage 3⁺ hosts induce *cERNI* within 3 h (4/4; not shown). Is the inducing ability of posterior cells due to FGF? In the presence of the FGF inhibitor SU5402, induction of both *ERNI* (0/5) and *Sox3* (0/5) by posterior cells is abolished.

Our findings suggest that neural induction is initiated before the

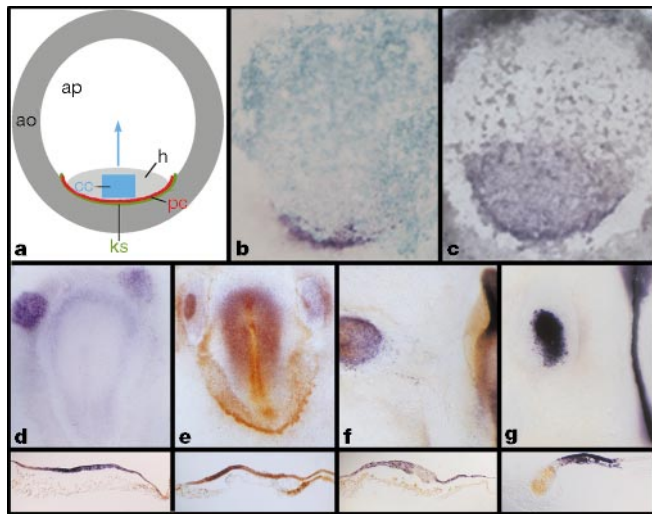


Figure 4 Organizer precursor cells initiate the neural induction cascade. **a**, Diagram showing the position of the two organizer precursor cell populations (pc and cc) at stage XI. ao, area opaca; ap, area pellucida; h, hypoblast; ks, Koller's sickle; pc, posterior cells; cc, central cells. The arrow indicates the movement of central cells occurring at stages XI–XIII. **b**, At stage X, *FGF8* is expressed in the posterior cells and subsequently (**c**; stage XII) in the forming hypoblast. *cERNI* (**d**) and *Sox3* (**e**) are strongly induced by posterior cells, whereas induction of *Sox2* (**f**) and *chordin* (**g**) requires signals from both organizer precursor populations. Insets below **d–g** show sections of the grafts in **d–g**, respectively. Quail embryos were used as hosts in **d, e** and as donors in **f, g** (in both cases quail cells labelled by QCPN, brown).

beginning of gastrulation by FGF emanating from a population of organizer precursors at the posterior margin of the embryo (perhaps reinforced by the spreading hypoblast). The coming together of this cell population with a second precursor population in the epiblast generates a fully functional organizer that provides the remaining signals in a cascade, including BMP antagonists. These results provide an explanation for the hitherto unexplained finding that in *Xenopus*, BMP antagonists do not induce neural tissue in the presence of dominant-negative FGF receptors^{18,19} and for controversial reports of FGF as a direct neural inducer^{20–23}. FGF signals are clearly not sufficient to generate a complete nervous system. But are they sufficient to sensitize the epiblast to BMP antagonists and to generate expression of later neural markers? The finding that *msx1* is upregulated by FGF8 raises the possibility that this is part of a mechanism that leads to self-maintaining activation of BMP signalling^{12,24}, which would be a required first step if the BMPs are later to be inhibited. On the other hand, neither FGF nor 5 h of signals from the node followed by BMP antagonists is sufficient to generate induction of *Sox2* or later neural markers^{4,12}. We propose that neural induction is a multi-step process of considerable complexity. FGF mimics the first 5 h of signals from the organizer, but further steps remain to be discovered. □

Methods

Differential screen

A quail node (stage 3⁺) was grafted into the area opaca of a chick host; after 5 h the epiblast that had been in contact with it was dissected ('induced epiblast'). A similar piece of tissue was isolated from the contralateral side of the same embryo ('uninduced epiblast'). From each tissue cDNA was generated from 10 different cell populations (10–15 cells) and libraries were constructed²⁵. One cDNA preparation was selected from each tissue; both had similar levels of the ubiquitously expressed ribosomal gene *S17*, whereas the early neural marker *Sox3* was present only in the induced tissue. The library from induced tissue was plated at low density and differentially screened using radioactively labelled total cDNA from both cell populations. Of 38 differentially expressed cDNAs, 36 corresponded to *cERNI*. A full-length clone (GenBank AF218814) was isolated by screening a chick gastrula-stage library. The full-length cDNA encodes a transcript of 4.5 kilobases (kb), which was confirmed by northern blot (not shown).

A quail primitive-streak-stage cDNA library was screened at low stringency to isolate a

quail *ERNI* homologue (GenBank AF218815). Zebrafish *ERNI* was isolated by genomic polymerase chain reaction (PCR) using degenerate primers, which amplified a 366-base pair (bp) fragment (*fERNI*).

Embryo culture and grafting

Fertile hens' eggs (White Leghorn, Spafas) and quails' eggs (Strickland) were incubated at 38 °C for 1–38 h to yield embryos at stages XI (ref. 26) to 10 (ref. 27). Host embryos (chick or quail) were maintained in New culture²⁸. Donor nodes from stage 3⁺/4 were transplanted to the inner margin of the area opaca of a host of the same stage. Organizer precursor cells (Fig. 4a) were mostly isolated from chick embryos where the fate maps are more complete^{10,17}. Posterior cells and associated Koller's sickle were dissected using syringe needles and transplanted to the area opaca of a quail host (stage 3⁺/4). To isolate central cells, hypoblast was removed and a square of epiblast (150 µm) excised.

FGF8-coated heparin beads were prepared as described¹². Chordin, Noggin, cerberus and BMP4-expressing cells were generated and grafted as described^{4,29}. Expression levels were routinely checked by immunostaining and immunoblots. To interfere with FGF signalling, pellets of 2,000–3,000 cells expressing a fusion protein of the extracellular domain of FGFR1 and rat IgG2a¹⁶ were grafted together with Hensen's node. AG1X2 beads were coated with 25 µM SU5402 (Calbiochem) in PBS for 1–2 h, washed and grafted together with the node or posterior cells into a host that had been pre-incubated briefly in inhibitor (2.5 µM).

Whole mount *in situ* hybridization and immunocytochemistry

Whole mount *in situ* hybridization and QCPN antibody staining were performed as described^{19,30}.

Received 14 March; accepted 3 May 2000.

- Harland, R. & Gerhart, J. Formation and function of Spemann's organizer. *Annu. Rev. Cell Dev. Biol.* **13**, 611–667 (1997).
- Wilson, P. A. & Hemmati-Brivanlou, A. Vertebrate neural induction: inducers, inhibitors, and a new synthesis. *Neuron* **18**, 699–710 (1997).
- Schier, A. F. & Talbot, W. S. The zebrafish organizer. *Curr. Opin. Genet. Dev.* **8**, 464–471 (1998).
- Streit, A. *et al.* Chordin regulates primitive streak development and the stability of induced neural cells, but is not sufficient for neural induction in the chick embryo. *Development* **125**, 507–519 (1998).
- Baker, J. C., Beddington, R. S. & Harland, R. M. Wnt signaling in *Xenopus* embryo inhibits BMP4 expression and activates neural development. *Genes Dev.* **13**, 3149–3159 (1999).
- Klingensmith, J., Ang, S. L., Bachiller, D. & Rossant, J. Neural induction and patterning in the mouse in the absence of the node and its derivatives. *Dev. Biol.* **216**, 535–549 (1999).
- Streit, A. & Stern, C. D. Neural induction. A bird's eye view. *Trends Genet.* **15**, 20–24 (1999).
- Gurdon, J. B. Embryonic induction—molecular prospects. *Development* **99**, 285–306 (1987).
- Streit, A. *et al.* Preventing the loss of competence for neural induction: HGF/SF, L5 and Sox-2. *Development* **124**, 119–1202 (1997).
- Hatada, Y. & Stern, C. D. A fate map of the epiblast of the early chick embryo. *Development* **120**, 2879–2889 (1994).
- Garcia-Martinez, V., Alvarez, I. S. & Schoenwolf, G. C. Locations of the ectodermal and nonectodermal subdivisions of the epiblast at stages 3 and 4 of avian gastrulation and neurulation. *J. Exp. Zool.* **267**, 431–446 (1993).
- Streit, A. & Stern, C. D. Establishment and maintenance of the border of the neural plate in the chick: involvement of FGF and BMP activity. *Mech. Dev.* **82**, 51–66 (1999).
- Chambers, D. & Mason, I. Expression of sprouty-2 during early development of the chick embryo is coincident with known sites of FGF signalling. *Mech. Dev.* **91**, 361–364 (2000).
- Storey, K. G. *et al.* Early posterior neural tissue is induced by FGF in the chick embryo. *Development* **125**, 473–484 (1998).
- Mohammadi, M. *et al.* Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science* **276**, 955–960 (1997).
- Ye, W., Shimamura, K., Rubenstein, J. L., Hynes, M. A. & Rosenthal, A. FGF and Shh signals control dopaminergic and serotonergic cell fate in the anterior neural plate. *Cell* **93**, 755–769 (1993).
- Izpisua-Belmonte, J. C., De Robertis, E. M., Storey, K. G. & Stern, C. D. The homeobox gene gooseoid and the origin of organizer cells in the early chick blastoderm. *Cell* **74**, 645–659 (1993).
- Launay, C., Fromentoux, V., Shi, D. L. & Boucaut, J. C. A truncated FGF receptor blocks neural induction by endogenous *Xenopus* inducers. *Development* **122**, 869–880 (1996).
- Sasai, Y., Lu, B., Piccolo, S. & De Robertis, E. M. Endoderm induction by the organizer-secreted factors chordin and noggin in *Xenopus* animal caps. *EMBO J.* **15**, 4547–4555 (1996).
- Amaya, E., Musci, T. J. & Kirschner, M. W. Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in *Xenopus* embryos. *Cell* **66**, 257–270 (1991).
- Lamb, T. M. & Harland, R. M. Fibroblast growth factor is a direct neural inducer, which combined with noggin generates anterior–posterior neural pattern. *Development* **121**, 3627–3636 (1995).
- Alvarez, I. S., Araujo, M. & Nieto, M. A. Neural induction in whole chick embryo cultures by FGF. *Dev. Biol.* **199**, 42–54 (1998).
- Hongo, I., Kengaku, M. & Okamoto, H. FGF signaling and the anterior neural induction in *Xenopus*. *Dev. Biol.* **216**, 561–581 (1999).
- Biehs, B., Francois, V. & Bier, E. The *Drosophila* short gastrulation gene prevents Dpp from autoactivating and suppressing neurogenesis in the neuroectoderm. *Genes Dev.* **10**, 2922–2934 (1996).
- Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).
- Eyal-Giladi, H. & Kochav, S. From cleavage to primitive streak formation: a complementary normal table and a new look at the first stages of the development of the chick. I. General morphology. *Dev. Biol.* **49**, 321–337 (1976).
- Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).
- Stern, C. D. & Ireland, G. W. An integrated experimental study of endoderm formation in avian embryos. *Anat. Embryol.* **163**, 245–263 (1981).

29. Streit, A. & Stern, C. D. Mesoderm patterning and somite formation during node regression: differential effects of chordin and noggin. *Mech. Dev.* **85**, 85–96 (1999).
30. Stern, C. D. Detection of multiple gene products simultaneously by in situ hybridization and immunohistochemistry in whole mounts of avian embryos. *Curr. Top. Dev. Biol.* **36**, 223–243 (1998).

Acknowledgements

We thank A. Rosenthal and W. Ye (Genentech) for the FGFR1-IgG construct; R. Lovell-Badge for *Sox3* and *Sox2* and T. Jessell for *S17*; C. Dulac for advice on the differential screen; B. Cigich for technical assistance; I. Skromne for Fig. 4b, c; C. Ang for zebrafish tissue; and T. Jessell, G. Sheng, K. Storey and D. Vasilias for helpful comments on the manuscript. Supported by the National Institute of Mental Health.

Correspondence and requests for materials should be addressed to C.D.S. (e-mail: cds20@columbia.edu).

Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2

Miyoko Higuchi*, Stefan Maas*†‡, Frank N. Single*†, Jochen Hartner*, Andrei Rozov§, Nail Burnashev§, Dirk Feldmeyer§, Rolf Sprengel* & Peter H. Seeburg*

* Departments of Molecular Neurobiology and § Cell Physiology, Max-Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany

† These authors contributed equally to this work

RNA editing by site-selective deamination of adenosine to inosine^{1,2} alters codons^{3,4} and splicing⁵ in nuclear transcripts⁶, and therefore protein function. ADAR2 (refs 7, 8) is a candidate mammalian editing enzyme that is widely expressed in brain and other tissues⁷, but its RNA substrates are unknown. Here we have studied ADAR2-mediated RNA editing by generating mice that are homozygous for a targeted functional null allele. Editing in *ADAR2*^{−/−} mice was substantially reduced at most of 25 positions in diverse transcripts^{3–6}; the mutant mice became prone to seizures and died young. The impaired phenotype appeared to result entirely from a single underedited position, as it reverted to normal when both alleles for the underedited transcript were substituted with alleles encoding the edited version exonic⁹. The critical position specifies an ion channel determinant¹⁰, the Q/R site^{3,6}, in AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionate) receptor¹⁰ GluR-B pre-messenger RNA. We conclude that this transcript is the physiologically most important substrate of ADAR2.

Mammalian transcripts that are known to be edited by site-selective adenosine deamination are expressed largely in brain: most encode subunits of ionotropic glutamate receptors (GluRs) that mediate fast excitatory neurotransmission^{3,10}. The only position edited to nearly 100% is the Q/R site of GluR-B, for which the mRNA contains an arginine (R) codon (CIG) in place of the genomic glutamine (Q) codon (CAG)³. The physiological importance of this codon substitution wrought by RNA editing was revealed by early onset epilepsy and premature death of mice heterozygous for an intron-11-modified *GluR-B*^{ΔECS} allele with Q/R site-uneditable transcripts^{11,12}.

Three mammalian adenosine deaminases acting on RNA

(ADAR1–ADAR3; refs 7, 8, 13) form a small family of candidate RNA-editing enzymes that operate on nuclear transcripts. Only ADAR2 edits the Q/R site in GluR-B pre-mRNA efficiently *in vitro*^{7,14,15}. Because ADAR2 is expressed in tissues other than brain⁷, distinct pre-mRNAs in different tissues may be substrates for this enzyme. To determine whether ADAR2 edits the Q/R site in GluR-B pre-mRNA *in vivo*, and to evaluate the general physiological significance of ADAR2-mediated RNA editing, we generated mice with functional null alleles for this enzyme.

A targeting vector for functional *ADAR2* gene ablation (Fig. 1a, b) was constructed by replacing most of exon 4 (ref. 16) with a *PGK-neo* gene; exon 4 encodes an essential adenosine deaminase motif^{3,7}. Chimaeric mice were generated by injection of a targeted embryonic stem (ES) cell clone¹⁷ into C57BL/6-derived blastocysts. *ADAR2*^{+/−} intercrosses produced *ADAR2*^{−/−} mice at Mendelian frequency, indicating that ADAR2 deficiency does not interfere with embryonic development. We found residual expression from the targeted *ADAR2*[−] allele through exon skipping, potentially leading to a truncated, enzymatically inactive ADAR2 form with intact RNA-binding domains (Fig. 1c). The expression of this truncated form might amount to less than 10% of ADAR2 in wild-type mice, as predicted from the severely reduced mutant transcript levels (mean ± s.d., 8 ± 3% of wild type, *n* = 3; postnatal day 14 (P14)) determined by ribonuclease (RNase) protection (Fig. 1d).

Heterozygous *ADAR2*^{+/−} mice were phenotypically normal, but *ADAR2*^{−/−} mice died between P0 and P20 and became progressively seizure-prone after P12, akin to *GluR-B*^{+/ΔECS} mice^{11,12}. Thus, we first studied the effect of ADAR2 deficiency on Q/R site editing of GluR-B pre-mRNA, the substrate for a nuclear RNA-dependent

Table 1 ADAR2 deficiency and site-selective adenosine deamination

Editing sites	ADAR2 ^{+/+}	ADAR2 ^{+/−}	ADAR2 ^{−/−}
GluR-B pre-mRNA			
Q/R*	98 ± 3	90 ± 3	10 ± 3
Hotspot1*†	50	45	60
Hotspot2*†‡			
+262	30	20	<10
+263	65	55	<10
+264	15	10	<5
GluR-B mRNA			
Q/R*	100 ± 1	100 ± 1	40 ± 4
AMPA-R† R/G§			
GluR-B	75	55	15
GluR-C	90	85	75
GluR-D	45	40	10
GluR5			
Q/R	64 ± 5	55 ± 2	40 ± 1
GluR6			
Q/R	86 ± 4	78 ± 4	29 ± 8
I/V	87 ± 2	79 ± 7	22 ± 5
Y/C	90 ± 4	82 ± 5	2 ± 1
5HT2C-R†			
A	75	70	70
B	80	75	30
C	15	10	<5
D	70	55	<5
ADAR2†‡			
−1	15	10	<10
+23	25	15	<10
+24	45	35	10

The editing sites have been described^{2–5,27}. Values, given as mean ± s.d. (*n* = 3), indicate the percentage of the edited version for the different editing sites analysed in the three genotypes. Values were obtained from whole-brain RNA by differential oligonucleotide-mediated hybridization of cloned RT-PCR products to distinguish between an adenosine (unedited) and a guanosine (edited) at the individual editing sites.

* Values from P14 mice with unmodified *GluR-B* alleles. All other values were from P40 mice with *GluR-B* alleles sequence modified at the Q/R site codon⁹.

† Values derived from the different peak heights for the two nucleotides in identical positions in DNA sequence chromatograms. Values from three mice were averaged and rounded to the nearest 5 or 0 position in each case.

‡ Sites +265 in hotspot2 and −2, +9 and +10 in ADAR2 were edited to <5% in wild type.

§ Flip and flop splice versions. As assessed from sequence chromatograms, the approximate representations of the flip forms were 50% for GluR-B, 75% for GluR-C and 65% for GluR-D; see ref. 27.

‡ Present address: Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, MA 02139, USA.

adenosine deaminase activity⁶. As determined from cloned polymerase chain reaction with reverse transcription (RT-PCR) products from brain RNA⁶, Q/R site editing in primary GluR-B transcripts was tenfold lower in *ADAR2*^{-/-} than in wild-type mice (10% compared with 98%, Table 1). This identifies ADAR2 as the principal RNA-editing enzyme at the Q/R site. The remaining low level of Q/R site editing in GluR-B pre-mRNA cannot be mediated by the residual, enzymatically inactive, truncated ADAR2 protein, but is mediated by another ADAR, perhaps ADAR1 (refs 14, 18), for which gene expression appeared unchanged in *ADAR2*^{-/-} mice (Fig. 1d).

The low extent of Q/R site editing of GluR-B pre-mRNA led to nuclear accumulation of incompletely processed primary GluR-B transcripts and to a fivefold reduction in GluR-B mRNA, as assessed by RNase protection (20 ± 5%; *n* = 3; Fig. 2a) and quantitative RT-PCR (20 ± 4% of wild type; *n* = 3; P14; not shown). The increased level of intron 11-containing GluR-B transcripts and the decrease in GluR-B mRNA were easily visualized by *in situ* hybridization (Fig. 2b). Editing is thus a prerequisite for efficient splicing and processing of the pre-mRNA. The edited GluR-B transcripts are preferentially spliced, as revealed by a shift in Q/R site editing from 10% to 40% when comparing intron-11-containing transcripts with GluR-B mRNA (Table 1). A defect in transcript processing caused by the interaction of the residual truncated ADAR2 protein with RNA can be excluded because GluR-B pre-mRNA accumulation is also

observed in *ADAR2*^{+/+} mice expressing the Q/R site-uneditable *GluR-B*^{ΔECS} allele¹¹.

The fivefold drop in mRNA reduced GluR-B subunit levels similarly (21 ± 3% of wild type; *n* = 3; P14), as indicated by western analysis of brain protein from *ADAR2*^{-/-} mice (Fig. 3a), and as visualized by immunocytochemistry in the hippocampal subfields (Fig. 3b). We also observed a moderate reduction in GluR-A protein (42 ± 5% of wild type; *n* = 3; Fig. 3a), barely detected by immunocytochemistry (Fig. 3b), with GluR-A mRNA being only slightly reduced (RNase protection: 83 ± 10% of wild type; *n* = 3; P14; Fig. 2a). The lower GluR-A protein levels may be a consequence of the severe reduction in GluR-B, the principal partner of GluR-A in AMPA receptor assembly⁹. Notably, mice lacking all GluR-A expression are phenotypically robust and normal in most respects¹⁹, precluding the possibility that the moderate GluR-A reduction contributes to the severe phenotype of *ADAR2*^{-/-} mice. NMDA receptor¹⁰ NR1 protein levels in *ADAR2*^{-/-} mice appeared unchanged (data not shown).

As predicted from previous studies^{11,12,20}, reduced expression and reduced Q/R site editing of GluR-B alters AMPA receptor channel properties mainly in principal neurones, which normally express fully edited GluR-B at levels sufficient for incorporation into most heteromeric AMPA receptors. Indeed, in accordance with the molecular analysis, AMPA receptor-mediated currents recorded in acute *ADAR2*^{-/-} brain slices from nucleated patches of CA1

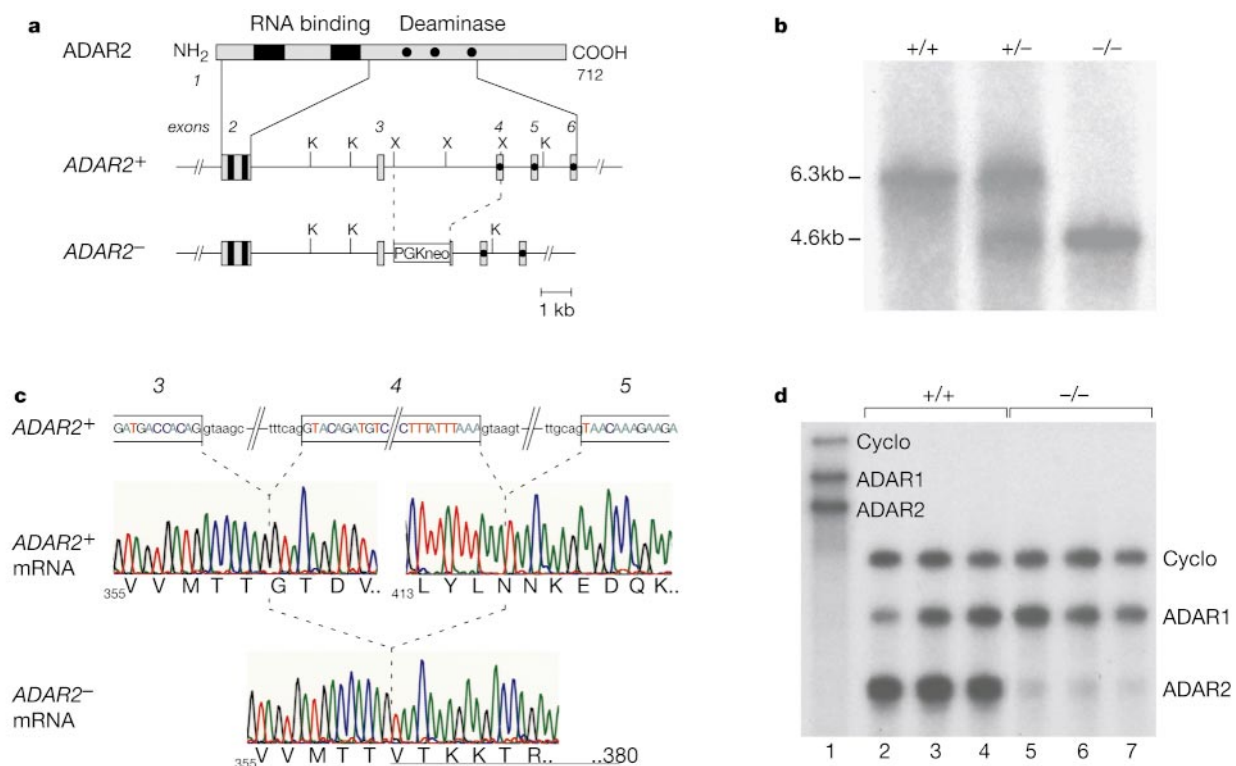


Figure 1 Targeted *ADAR2* allele (Mouse Genome Database symbol, *Adarb1*) and transcript. **a**, Domain structure of the RNA-dependent adenosine deaminase ADAR2 and exon-intron organization of *ADAR2*⁺ and *ADAR2*⁻ alleles. Protein domains^{3,7} are connected to their corresponding exons¹⁶. In the *ADAR2*⁻ allele, most of intron 3 and exon 4 is replaced by *PGK-neo* (dashed lines). Filled boxes, double-stranded RNA-binding domains in protein and exons; filled circles, conserved sequences in deaminase domain of ADARs³. X, *XhoI*; K, *KpnI*. **b**, Southern analysis of *ADAR2* alleles in *KpnI*-restricted genomic DNA from mouse liver with a cDNA probe encoding residues 323–452 from exons 3–5. +/+, wild type; +/-, heterozygote; -/-, *ADAR2*-deficient homozygote. **c**, DNA sequence analysis of RT-PCR products from *ADAR2*⁺ and *ADAR2*⁻ mRNAs. A segment with exons 3–5 of the *ADAR2*⁻ allele depicts nucleotides coloured as in chromatograms for wild-type

and mutant cDNA sequences. Amino acids are given as single letters, numbered according to ADAR2. Exon sequence deleted in *ADAR2*⁻ is indicated by stippled lines. Note the frameshift caused by exon 4 deletion (underlined grey residues). **d**, RNase protection with whole-brain RNA by mRNAs for cyclophilin (Cyclo) as internal standard, ADAR1, and ADAR2 in three wild-type (+/+) and three *ADAR2*-deficient (-/-) mice. Lane 1, unprotected radiolabelled antisense RNA probes for cyclophilin (324 nt), ADAR1 (283 nt) and ADAR2 (248 nt). Lanes 2–7, protected probe fragments for cyclophilin (244 nt), ADAR1 (203 nt, exon 6–7 sequences²⁵ encoding RNA-binding domains) and ADAR2 (168 nt, exons 2–3 sequences¹⁶ for RNA-binding domains). A probe derived from ADAR2 exon 6–7 sequences¹⁶ for part of the deaminase domain yielded the same reduction for *ADAR2*⁻ allele expression (not shown).

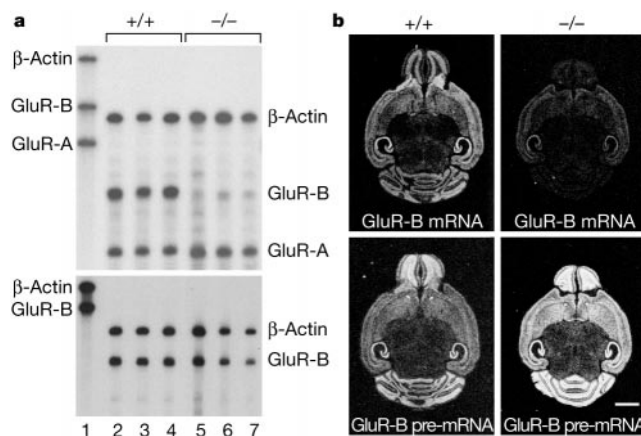


Figure 2 GluR-B and GluR-A transcript analyses. **a**, RNase protection. Lane 1, unprotected radiolabelled antisense RNA probes for β-actin as internal standard (334 nt), GluR-B²⁶ (upper gel, 270 nt; lower gel, 272 nt) and GluR-A (221 nt); lanes 2–4, protection by whole-brain RNA from P14 *ADAR2*^{+/+} mice (+/+); lanes 5–7, from P14 *ADAR2*^{-/-} mice (-/-). Protected fragment sizes in nt: β-actin, 245; GluR-B, 190 (upper gel, exons 11–12) and 192 (lower gel, exons 3–4); GluR-A, 141 (exons 3–4). Note that the reduction in

GluR-B mRNA in the mutant is only revealed with the exon 11–12 probe (upper gel), but not with the exon 3–4 probe (sum of pre- and mRNA; lower gel). **b**, *In situ* hybridization. Horizontal brain sections of *ADAR2*^{+/+} and *ADAR2*^{-/-} mice were hybridized with ³⁵S-labelled oligonucleotide probes²⁴ specific for GluR-B mRNA (36 nt, spanning exons 11–12) and pre-mRNA (37 nt, intron 11). Note the substantially higher levels of GluR-B pre-mRNA in the mutant and of GluR-B mRNA in wild type. Scale bar, 2 mm.

pyramidal cells and other principal neurones (cerebellar Purkinje cells and neocortical layer 5 pyramidal neurones) displayed, relative to wild type, pronounced rectification, increased desensitization rates and 30-fold higher Ca²⁺ permeability (P_{Ca}/P_{Na} 3.34–3.94 ($n = 4–8$) compared with 0.12–0.14 ($n = 4–10$)). Moreover, the macroscopic AMPA receptor-mediated conductance in CA1 pyramidal cells of *ADAR2*^{-/-} mice was significantly higher (40.8 ± 4.0 ($n = 10$) compared with $21.4 \pm 3.0 \mu S$ ($n = 3$) in wild type) than in our mouse mutants expressing editing-deficient *GluR-B* genes^{11,12}. As reported¹², the increase in macroscopic AMPA conductance, caused by the higher single-channel conductance in absence of GluR-B(R) participation¹⁰, appears to be the primary cause for the high seizure susceptibility of young mice expressing Q/R site-unedited GluR-B transcripts.

To show that the prematurely lethal phenotype of *ADAR2*^{-/-} mice

is engendered by insufficient Q/R site editing of GluR-B, we attempted a rescue of this phenotype by introducing a *GluR-B*^R allele⁹. This allele, generated by targeting, contains an arginine (R) codon for the Q/R site, and its expression in *ADAR2*^{-/-} mice is independent of Q/R site editing. We found that the severely compromised phenotype of *ADAR2*^{-/-} mice was partially rescued already in heterozygous *ADAR2*^{-/-}/*GluR-B*^{R/R} mutants, which survived up to P35 (Fig. 4). Moreover, their phenotype was less impaired than that of *GluR-B*^{+/^{ΔECS} mice, consistent with the hypothesis that phenotypic impairment increases with the Q/R ratio of GluR-B¹². In both *ADAR2*^{-/-}/*GluR-B*^{+/^R and *GluR-B*^{+/^{ΔECS} mutants, one GluR-B allele expresses exclusively GluR-B(R), but primary transcripts from the second allele are underedited and the corresponding mRNA levels are reduced. In mRNA derived from the *GluR-B*^{ΔECS} allele the Q/R site is not edited at all, whereas mRNA from the *GluR-B*^R allele is Q/R site edited to 40% in the *ADAR2*^{-/-} background. Thus, the Q/R ratio in GluR-B-containing AMPA receptors is lower in *ADAR2*^{-/-}/*GluR-B*^{+/^R than in *GluR-B*^{+/^{ΔECS}}}}}}

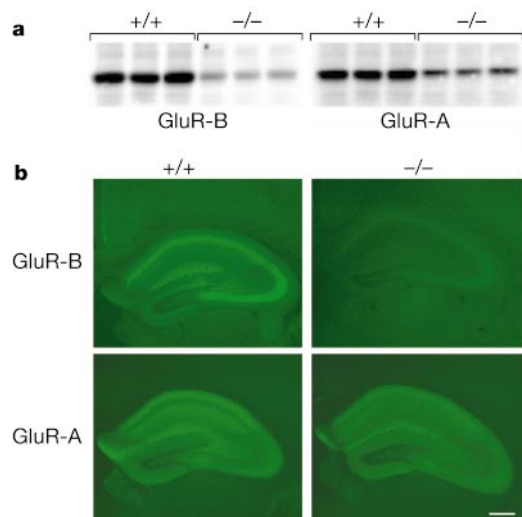


Figure 3 GluR-B and GluR-A protein analyses. **a**, Western blots of brain protein from three *ADAR2*^{+/+} (+/+) and three *ADAR2*^{-/-} (-/-) mice with antibodies for two AMPA receptor subunits. **b**, Immunocytochemistry showing blow-ups of the hippocampus in coronal vibratome sections (40 μm) of P14 *ADAR2*^{+/+} and *ADAR2*^{-/-} brains, probed with antibodies to GluR-B (upper panels) and GluR-A (lower panels) and stained with FITC-conjugated secondary antibody. Note that a moderate GluR-A reduction is apparent in **a**, but barely detectable in **b**. Scale bar, 0.3 mm.

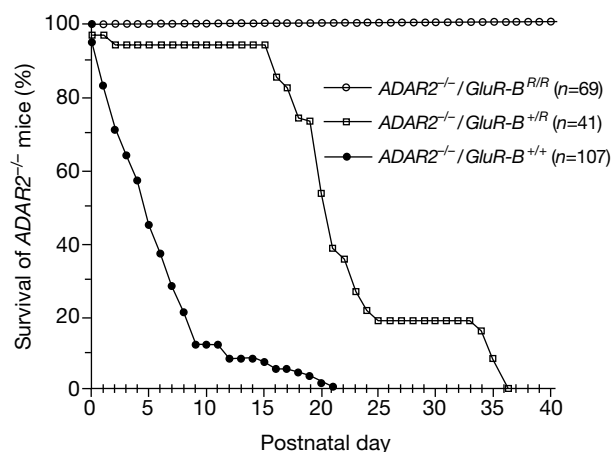


Figure 4 Survival of *ADAR2*^{-/-} mice. Data points depict the percentage of mice alive at the indicated postnatal days. The number of mice observed is listed in parentheses for each genotype. Note that the life span of *ADAR2*-deficient mice increased with the copy number of the sequence-modified *GluR-B*^R allele, entirely consistent with the reduction in Q/R ratio, as elucidated by previous studies on the effect of Q/R site-uneditable *GluR-B* alleles in targeted mouse mutants^{11,12}.

mice, and the phenotype is less affected. Full rescue of the impaired $ADAR2^{-/-}$ phenotype was achieved after introducing the second $GluR-B^R$ allele. $ADAR2^{-/-}/GluR-B^{R/R}$ mice appeared normal at all ages, like $GluR-B^{R/R}$ mice⁹, as judged by food intake, weight increase, postnatal development, breeding and general behaviour. This identifies the GluR-B transcript as the most critical substrate for ADAR2 in the mammal.

We investigated whether RNA editing at other known sites is affected by functional ADAR2 ablation. We observed reduced editing at the majority of 25 positions in diverse transcripts (Table 1); even hemizyosity for ADAR2 led at several sites to a modest decline in the extent of editing (Table 1). Notably, the actual reduction in the extent of editing may not have been revealed, as many of the sites were analysed at the mRNA rather than the pre-mRNA stage. Nevertheless, the results indicate a pleiotropic action of ADAR2, not entirely anticipated from *in vitro* studies^{14,21}; ADAR2 may act in concert with other RNA-editing enzymes, perhaps assembled with additional RNA interacting proteins in a multi-component nuclear 'editosome'. The reduced editing at many sites might also reflect, at least in part, interference with other ADARs by the residual truncated ADAR2 protein (Fig. 1c); if so, affected editing levels might not indicate true catalysis by ADAR2. The severe reduction in transcript levels from the mutant allele (Fig. 1d), however, may render such interference unlikely.

In summary, we have shown that RNA editing can control the rate of splicing in nuclear transcripts; a link between editing and splicing has also been reported in *Drosophila*²². Moreover, our study identified the Q/R site codon in GluR-B pre-mRNA as a critical physiological substrate of ADAR2, apparently the only one at which editing is essential for viability, at least during the first few weeks of life. Whether or not dependence of viability on Q/R site editing extends into adulthood will be investigated by conditional ADAR2 gene expression. As $ADAR2^{-/-}$ mice were phenotypically rescued by $GluR-B^R$ alleles, the effects caused by underediting of other ADAR2 substrates need to be elucidated by future analyses of $ADAR2^{-/-}/GluR-B^{R/R}$ mice. □

Methods

Extent of editing

RT-PCR products from brain RNA of the different genotypes were generated with primers specific for the pre-mRNAs or mRNAs (details of oligonucleotides are available from the author on request). Amplicons were cloned directionally into M13 RF DNA and more than 1,000 recombinant single-stranded phage plaques were hybridized to oligonucleotide probes that distinguish between edited and unedited positions in the different sequences⁶. Alternatively, amplicons were sequenced directly, and the approximate extent of editing was derived from the different peak heights for the two nucleotides occurring in identical positions in DNA sequence chromatograms.

GluR-B mRNA levels by RT-PCR

RT-PCR was used with primers to co-amplify all four AMPA receptor subunit cDNAs²³. Amplicons were cloned directionally, and filters containing more than 1,000 recombinant phage plaques were probed sequentially for the different subunit cDNAs. The numeric representations (in percent) of the four subunit sequences were for $ADAR2^{-/-}$ mice: GluR-A, 57; GluR-B, 14; GluR-C, 20; GluR-D, 9; for $ADAR2^{+/+}$ mice 36, 41, 15, 8, and for wild-type mice 39, 44, 11, 6.

RNase protection

RNase protection (RPAIII Ribonuclease protection assay kit, Ambion) was used to determine transcripts levels for GluR-A, GluR-B, ADAR1 and ADAR2. Individual reactions were performed with 20 µg total whole-brain RNA, and radiolabelled antisense RNA probes were as described in Fig. 2.

In situ hybridization

In situ hybridization was performed on horizontal cryostat sections of P14 mouse brains with 3'-end-labelled oligonucleotides²⁴. Exposure to X-ray film was for 7 days. Control sections hybridized with labelled probe in presence of a 100-fold excess of the unlabelled oligonucleotide generated no signal.

Western analysis and immunocytochemistry

Western blots with 10 µg of P14 whole-brain protein extracts per lane were successively

probed, after intermittent stripping, with antibodies to GluR-B (1:200; Pharmingen) and GluR-A (1:300; Chemicon); secondary antibody was conjugated to horseradish peroxidase (1:40,000; Amersham) and detection was with Super Signal West (Pierce). P14 brain sections were processed for immunocytochemistry⁹, incubated with antibodies to GluR-B (1:10) and GluR-A (1:50) and stained with FITC-conjugated goat anti-rabbit IgG (1:100; Jackson ImmunoResearch).

Electrophysiology

AMPA receptor-mediated currents in whole-soma 'nucleated' patches of identified neurones in acute brain slices of P14 mice were elicited and analysed as described¹².

Received 10 March; accepted 22 May 2000.

- Bass, B. L. RNA editing. An I for editing. *Curr. Biol.* **5**, 598–600 (1995).
- Rueter, S. & Emeson, R. in *Modification and Editing of RNA* (eds Grosjean, H. & Benne, R.) 343–361 (ASM, Washington DC, 1998).
- Seeburg, P. H., Higuchi, M. & Sprengel, R. RNA editing of brain glutamate receptor channels: mechanism and physiology. *Brain Res. Rev.* **26**, 217–229 (1998).
- Burns, C. M. *et al.* Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387**, 303–308 (1997).
- Rueter, S. M., Dawson, T. R. & Emeson, R. B. Regulation of alternative splicing by RNA editing. *Nature* **399**, 75–79 (1999).
- Higuchi, M. *et al.* RNA editing of AMPA receptor subunit GluR-B: A base-paired intron–exon structure determines position and efficiency. *Cell* **75**, 1361–1370 (1993).
- Melcher, T. *et al.* A mammalian RNA editing enzyme. *Nature* **379**, 460–464 (1996).
- Bass, B. L. *et al.* A standardized nomenclature for adenosine deaminases that act on RNA. *RNA* **3**, 947–949 (1997).
- Kask, K. *et al.* The AMPA receptor subunit GluR-B in its Q/R site-unedited form is not essential for brain development and function. *Proc. Natl Acad. Sci. USA* **95**, 13777–13782 (1998).
- Dingledine, R., Borges, K., Bowie, D. & Traynelis, S. F. The glutamate receptor ion channels. *Pharmacol. Rev.* **51**, 7–61 (1999).
- Brusa, R. *et al.* Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. *Science* **270**, 1677–1680 (1995).
- Feldmeyer, D. *et al.* Neurological dysfunctions in mice expressing different levels of the Q/R site-unedited AMPAR subunit GluR-B. *Nature Neurosci.* **2**, 57–64 (1999).
- Melcher, T. *et al.* RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J. Biol. Chem.* **271**, 31795–31798 (1996).
- Maas, S. *et al.* Different structural and enzymatic requirements for RNA editing in glutamate receptor pre-mRNAs. *J. Biol. Chem.* **271**, 12221–12226 (1996).
- Lai, F., Chen, C. X., Carter, K. C. & Nishikura, K. Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. *Mol. Cell. Biol.* **17**, 2413–2424 (1997).
- Villard, L., Tassone, F., Haymowicz, M., Welborn, R. & Gardiner, K. Map location, genomic organization and expression patterns of the human RED1 RNA editase. *Somat. Cell. Mol. Gen.* **23**, 135–145 (1997).
- Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
- Dabiri, G. A., Lai, F., Drakas, R. A. & Nishikura, K. Editing of GluR-B ion channel RNA in vitro by recombinant double-stranded RNA adenosine deaminase. *EMBO J.* **15**, 34–45 (1996).
- Zamanillo, D. *et al.* Importance of AMPA receptors for hippocampal synaptic plasticity but not for spatial learning. *Science* **284**, 1805–1811 (1999).
- Jia, Z. *et al.* Enhanced LTP in mice deficient in the AMPA receptor GluR2. *Neuron* **17**, 945–956 (1996).
- Herb, A., Higuchi, M., Sprengel, R. & Seeburg, P. H. Q/R site editing in kainate receptor GluR5 and GluR6 pre-mRNAs requires distant intronic sequences. *Proc. Natl Acad. Sci. USA* **93**, 1875–880 (1996).
- Reenan, R. A., Hanrahan, C. J. & Ganetzky, B. The *mle* RNA helicase mutation in *Drosophila* results in a splicing catastrophe of the *para* Na⁺ channel transcript in a region of RNA editing. *Neuron* **25**, 139–149 (2000).
- Geiger, J. R. P. *et al.* Relative abundance of subunit mRNAs determines gating and Ca²⁺ permeability of AMPA receptors in principal neurons and interneurons in rat CNS. *Neuron* **15**, 193–204 (1995).
- Wisden, W. & Morris, B. J. in *In Situ Protocols for the Brain* (eds Wisden, W. & Morris, B. J. 9–30 (Academic, London, 1994).
- Wang, Y., Zeng, Y., Murray, J. M. & Nishikura, K. Genomic organization and chromosomal localization of the human dsRNA adenosine deaminase gene: the enzyme for glutamate-activated ion channel RNA editing. *J. Mol. Biol.* **254**, 184–195 (1995).
- Köhler, M., Kornau, H. -C. & Seeburg, P. H. The gene for the principal AMPA receptor subunit GluR-B: organization and sequences for alternatively spliced and edited transcripts. *J. Biol. Chem.* **269**, 17367–17370 (1994).
- Lomeli, H. *et al.* Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **266**, 1709–1713 (1994).

Acknowledgements

We thank R. Wenthold for the antibody to GluR-B; A. Nagy for the murine embryonic stem cell line R1; K. Kask for help with $GluR-B^R$ mice; F. Zimmermann for blastocyst injection; S. Grünwald and H. Grosskurth for DNA sequencing; U. Amtmann for *in situ* hybridization; and H. Avci, C. Faul and C. Baust for technical help. This work was supported, in part, by the Deutsche Forschungsgemeinschaft, the Human Frontier Science Program and the Bristol-Myers Squibb foundation.

Correspondence and requests for materials should be addressed to P.H.S. (e-mail: seeburg@mpimf-heidelberg.mpg.de).

Therapeutic haemoglobin synthesis in β -thalassaemic mice expressing lentivirus-encoded human β -globin

Chad May^{*†‡}, Stefano Rivella^{*}, John Callegari^{*}, Glenn Heller[§], Karen M. L. Gaensler^{||}, Lucio Luzzatto^{*¶} & Michel Sadelain^{*†‡¶}

^{*} Department of Human Genetics, [†] Immunology Program, and Departments of [§] Epidemiology and Biostatistics, [¶] Medicine and [#] Pediatrics, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA
[‡] Weill Graduate School of Medical Sciences, Cornell University, New York, New York 10021, USA
^{||} Department of Medicine, University of California, San Francisco, California 94143, USA

The stable introduction of a functional β -globin gene in haematopoietic stem cells could be a powerful approach to treat β -thalassaemia¹ and sickle-cell disease². Genetic approaches aiming to increase normal β -globin expression in the progeny of autologous haematopoietic stem cells³ might circumvent the limitations and risks of allogeneic cell transplants⁴. However, low-level expression, position effects and transcriptional silencing hampered the effectiveness of viral transduction of the human β -globin gene when it was linked to minimal regulatory sequences⁵. Here we show that the use of recombinant lentiviruses enables efficient transfer and faithful integration of the human β -globin gene together with large segments of its locus control region. In long-term recipients of unselected transduced bone marrow cells, tetramers of two murine α -globin and two human β -globin molecules account for up to 13% of total haemoglobin in mature red cells of normal mice. In β -thalassaemic heterozygous mice higher percentages are obtained (17% to 24%), which are sufficient to ameliorate anaemia and red cell morphology. Such levels should be of therapeutic benefit in patients with severe defects in haemoglobin production.

Onco-retroviral-mediated transfer in mouse haematopoietic stem cells (HSCs) permits erythroid-specific expression of the human β -globin gene, but expression is low and limited by chromosomal position effects⁵. The same outcome prevails in transgenic mice unless genomic elements encompassing DNase I hypersensitive sites (HS) located 60 kilobases (kb) upstream of the human β -globin gene^{6,7}, referred to as the locus control region (LCR), are linked to the globin gene⁸. Incorporation of small elements spanning DNase HS2, HS3 and HS4 into viral vectors increases β -globin expression in mouse erythroleukaemia (MEL) cells^{9,10}. However, low-level expression, strong position effects and transcriptional inactivation are still observed in bone marrow chimaeras^{5,11}. Studies in transgenic mice¹² and deletional analyses¹³ support the view that coordinated interaction of several genetic elements including the LCR is required for physiologic β -globin gene expression^{12–15}. We therefore thought that incorporation of large elements spanning HS2, HS3 and HS4^{16–18} in a vector might enhance β -globin expression beyond the levels previously achieved using arrayed minimal core elements^{5,9–11}, and thus might diminish position effects and vector silencing. The efficient transduction of large genomic fragments using onco-retroviral vectors has proved to be severely curtailed by splicing and other alterations affecting the stability of the recombinant genomes^{9,10,16}. Here we report how these problems may be overcome by using vectors derived from human immunodeficiency virus 1, a retrovirus that has the ability to regulate packaging of unspliced viral genomes¹⁹.

We constructed two recombinant lentiviruses carrying β -globin transcription units (Fig. 1a, b). RNS1 contains a minimal LCR comprising previously tested core elements of HS2, HS3 and HS4

(ref. 9). To create TNS9, large fragments encompassing HS2, HS3 and HS4 were introduced instead of the corresponding core elements. The genomic integrity of each recombinant lentivirus was assessed after transfection and reverse transcription. The dominant RNA species for RNS1 and TNS9 was the full-length transcript (Fig. 1c), suggesting that the intact recombinant genome was available for packaging. This was confirmed by the presence of single, intact proviral structures in transduced cells (Fig. 1d).

To investigate the tissue specificity, stage specificity and expression level of the vector-encoded human β -globin gene, we transduced RNS1 and TNS9 into MEL cells, lymphoid Jurkat cells and myeloid HL-60 cells. To induce β -globin transcription, transduced MEL cell pools were differentiated using hexamethylene bisacetamide (HMB). Human β -globin (β^A) and mouse β -globin transcripts were measured by quantitative primer extension. After normalization to vector copy number and to endogenous β -globin expression per allele, human β -globin levels were $14.2 \pm 4.7\%$ for RNS1 and $71.3 \pm 2.3\%$ for TNS9 in pooled MEL cells (Fig. 2a). No human β -globin expression was detected in non-induced MEL, Jurkat and HL-60 cells (Fig. 2a), indicating that human β -globin expression was appropriately regulated in terms of tissue specificity and state of differentiation. We generated a panel of MEL cell clones that carried a single copy of either vector to distinguish whether the increased expression obtained in HMB-treated MEL cells transduced with TNS9 rather than RNS1 was the result of an increase in β^A

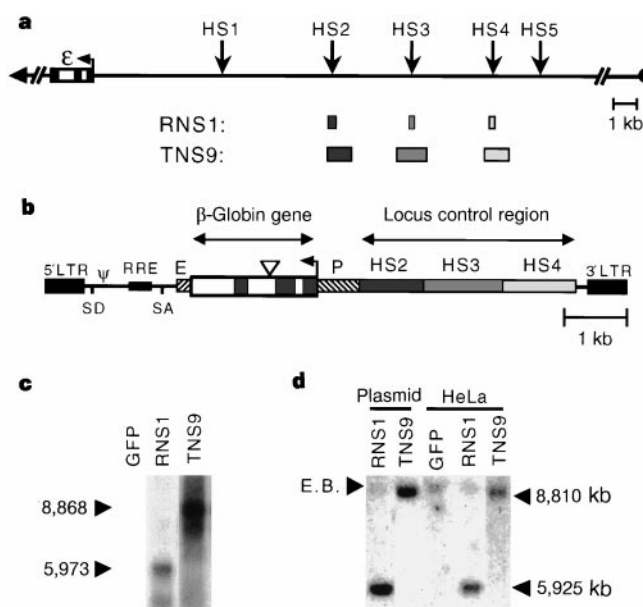


Figure 1 Structure and stability of recombinant lentiviral vectors. **a**, Genomic structure of the human β -globin LCR and mapping of fragments introduced in RNS1 and TNS9. Main DNase I hypersensitive sites (HS) are indicated. RNS1 contains a 1.0-kb LCR consisting of the 423-bp HS2, 280-bp HS3 and 283-bp HS4 core elements previously studied in an MLV based vector^{9,9}. TNS9 was generated by replacing the core HS2 element of RNS1 with an 840-bp HS2 fragment, the core HS3 element with a 1,308-bp HS3 fragment, and the core HS4 element with a 1,069-bp HS4 fragment. Boxes represent HS fragments drawn to scale. **b**, Representation of TNS9. Exons and introns of the human β -globin gene (including the IVS2 deletion⁹) are represented by filled and open boxes, respectively. Splice donor (SD), splice acceptor (SA), packaging region (Ψ), rev-response element (RRE), human β -globin promoter (P) and 3' β -globin enhancer (E) are indicated. **c**, Northern blot analysis shows full-length RNA transcripts indicating that the recombinant lentiviral genomes are stable. A lentiviral vector encoding GFP was used as control. **d**, Southern blot analysis on genomic DNA from transduced cells, digested once in each long terminal repeat (LTR), results in a single band corresponding to the expected size for both vectors, indicating that the proviral structure is not rearranged. Plasmid controls are shown in the left panel. The 9.0-kb endogenous murine band (E.B.) is indicated.

expression per cell or of an increase in the fraction of cells expressing human β -globin. Transduced MEL cells were subcloned by limiting dilution immediately after transduction, avoiding any bias towards favourable chromosomal integration sites as produced by drug selection⁵. The proportion of clones expressing human β -globin varied significantly between the two vectors (Fig. 2b). One out of ten RNS1 positive clones yielded measurable human β -globin transcripts, in contrast to 12 out of 12 for TNS9 ($P < 0.01$, Fisher's exact test). Cells bearing TNS9 also expressed higher levels of human β -globin than did those bearing RNS1 ($P < 0.01$, Wilcoxon rank sum test). These findings established that both the level and probability of expression at random integration sites was increased with the TNS9 vector.

To investigate the function of these vectors *in vivo*, we transduced and transplanted murine bone marrow cells without any selection in syngeneic, lethally irradiated recipient mice. The average vector copy number in peripheral blood cells, measured periodically for 24 weeks (Fig. 3a), showed highly efficient gene transfer with both vectors (1.8 ± 0.6 and 0.8 ± 0.6 average vector copies per cell for

RNS1 and TNS9, respectively, at week 24). For RNS1, human β -globin transcripts were in the range of $7.0 \pm 3.0\%$ after 6 weeks, later decreasing to 2.1% and 3.4% at week 16 and 24, respectively (expressed as fraction of total human + mouse β -globin transcripts, Fig. 3b). In contrast, cells transduced with TNS9 maintained human

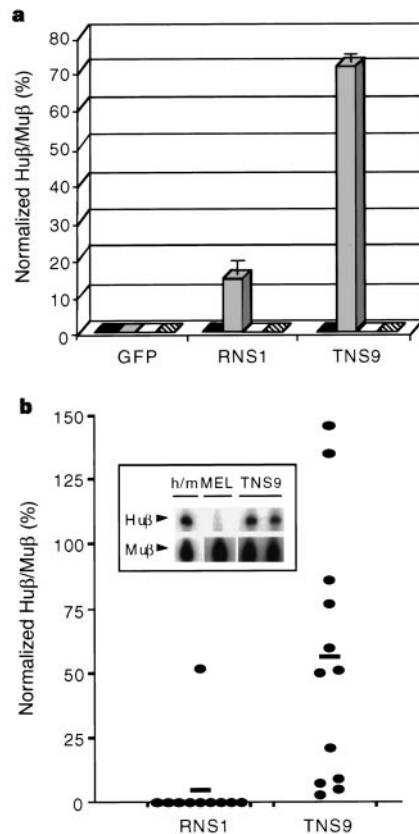


Figure 2 Increased mean β -globin expression in MEL cells and proportion of clones expressing detectable human β -globin with the TNS9 vector. **a**, MEL, Jurkat and HL-60 cells were transduced with RNS1, TNS9 or control GFP recombinant lentivirus. Human β -globin RNA expression in HMBa-induced MEL cells (grey bars) was measured by quantitative primer extension and normalized to mouse β -globin RNA expression per locus. Expression was then normalized to the vector copy number determined by Southern blot. No human β -globin RNA expression was detected in non-induced MEL (black bars), Jurkat (white bars) or HL-60 cells (hatched bars), indicating that globin expression was erythroid- and differentiation-specific. **b**, MEL cell clones carrying an intact single copy of RNS1 or TNS9 were identified and induced to terminal maturation to study variability of expression at individual vector integration sites. Ten RNS1 and twelve TNS9 clones were analysed. The expression of human β -globin relative to mouse β -globin per locus is shown. Black bar represents mean expression levels of each vector. Inset, representative primer extension reactions of two TNS9 transduced MEL cell clones, non-transduced MEL cells and a positive control (equal mix of human and mouse blood RNA, h/m).

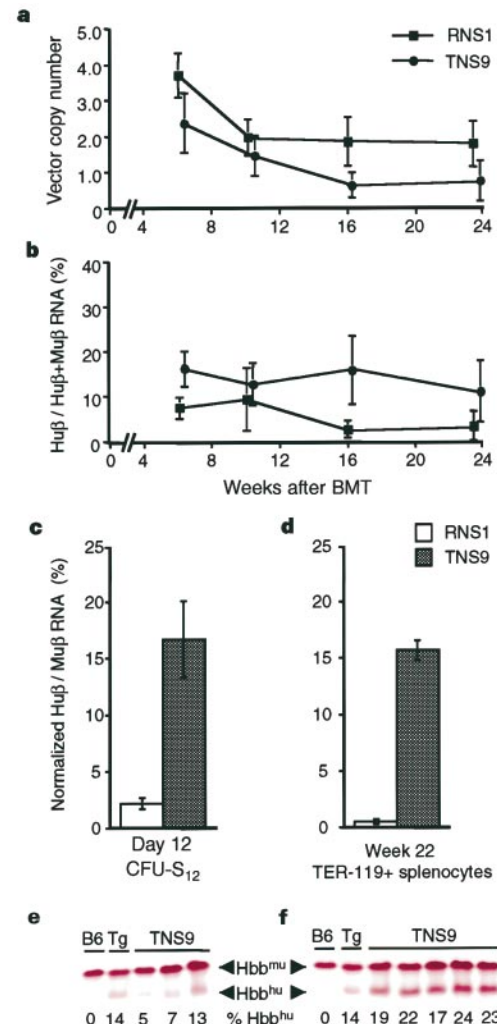


Figure 3 Long-term stability of vector copy number, human β -globin RNA levels and haemoglobin tetramers in peripheral blood of bone marrow chimaeras. Vector copy number and human β -globin RNA transcripts were measured during a 24-week period in mice transplanted with RNS1 ($n = 8$) or TNS9 ($n = 10$) transduced bone marrow. **a**, Vector copy number was assessed by Southern blot analysis of genomic DNA isolated from peripheral blood at weeks 6, 10, 16 and 24. **b**, Human β -globin RNA levels in peripheral blood are expressed relative to total human and mouse β -globin RNA (Hu β /Hu β + Mu β). **c, d**, Human β -globin expression per vector copy in spleen cells 12 days (**c**, pooled CFU-S₁₂) and 22 weeks (**d**, TER-119+ cells) after bone marrow transplantation. Human β -globin RNA expression is measured relative to the level of mouse β -globin RNA expression in the splenocytes of bone marrow chimaeras transduced with RNS1 ($n = 6$) or TNS9 ($n = 4$). Mean RNA expression (per vector copy) is normalized to endogenous gene expression. Vector copy number was comparable but no human β -globin RNA was detected in the thymus of the same animals, (data not shown). **e**, Hbb^{hu} (mouse α_2 : human β_2^A) levels were analysed 24 weeks after transplantation using cellulose acetate gel electrophoresis. Haemoglobin samples are from a control adult C57BL/6 mouse (B6), the transgenic human β -globin YAC mouse line A85.68 (Tg; ref. 20), and three TNS9 bone marrow chimaeras. Fraction of Hbb^{hu} relative to total Hbb (Hbb^{hu} / Hbb^{hu} + Hbb^{mu}) is indicated below each sample (see Table 1 in Supplementary Information for detailed analyses). **f**, Hbb^{hu} formation in mice transplanted with β^0 heterozygote (Hbb^{th3/+}) bone marrow cells transduced with TNS9 ($n = 5$), and analysed 8 weeks after transplantation.

β -globin transcript levels in the 10–20% range during the same time period. To assess transcriptional activity per vector copy, steady-state RNA transcripts and vector copy number were quantified in pooled CFU-S₁₂ and in erythroid TER-119+ spleen cells (Fig. 3c, d). Twelve days after transplantation, human β -globin expression per vector copy was $2.3 \pm 0.5\%$ and $16.8 \pm 3.9\%$ for RNS1 and TNS9, respectively (normalized to murine β -globin expression per endogenous allele, Fig. 3c). Twenty weeks later, these values were $0.5 \pm 0.1\%$ (significantly lower than on day 12, $P = 0.02$) and $15.8 \pm 0.9\%$, respectively (Fig. 3d). These findings established that the larger LCR fragments increased globin expression *in vivo* and, furthermore, suggested that TNS9 is more resistant to transcriptional silencing than is RNS1.

The levels of TNS9-encoded human β -globin RNA measured in peripheral blood suggested that therapeutically relevant levels of human β -globin could be produced. Haemoglobin tetramers incorporating vector-encoded human β^A and endogenous murine α -globin (designated Hbb^{hu}) were quantitated in peripheral blood red cell lysates after cellulose acetate gel fractionation. Hbb^{hu} levels accounting for up to 13% of total haemoglobin were found 24 weeks after transplantation (Fig. 3e, Table 1 in Supplementary Information). In the same assays, transgenic mice bearing one copy of a 230-kb yeast artificial chromosome (YAC) encompassing the entire human β -globin-like gene cluster²⁰ showed 14% of their total haemoglobin incorporating human β^A (Fig. 3e). No haemoglobin tetramers containing human β^A were measurable in any of the mice bearing RNS1 (Table 1 in Supplementary Information). The proportion of mature peripheral blood red cells expressing human β^A was elevated in most TNS9 bone marrow chimaeras, as shown by dual staining of human β^A and TER-119 (Table 1 in Supplementary Information). In contrast, chimaeras engrafted with RNS1-transduced bone marrow showed highly variable fractions of weakly staining β^A -positive erythrocytes (Table 1 in Supplementary Information). Normalized to the fraction of circulating β^A -positive mature red cells, the levels of haemoglobin containing lentivirus-encoded β^A were on average 64% of those obtained in the YAC transgenic mice (Table 1 in Supplementary Information).

To ascertain that true HSCs were transduced, we carried out secondary transplants using marrow from primary recipients collected 24 weeks after transplantation. The TNS9 and RNS1 vectors were readily detected in all secondary recipients 13 weeks after transplantation. Human β -globin expression was maintained in all recipients of TNS9-transduced marrow (Table 2 in Supplementary Information). The successful transduction of HSCs was confirmed by integration site analyses (Fig. 4).

In view of the high levels of expression, we tested the extent to which the TNS9 vector could correct the phenotype of thalassaemic

cells using β^0 -thalassaemic heterozygote mice that lack a copy of their b1 and b2 β -globin genes (Hbb^{th3/+})²¹. These heterozygotes have a clinical phenotype similar to human thalassaemia intermedia and exhibit chronic anaemia (haematocrit 28–30%, haemoglobin 8–9 g dl⁻¹) and anomalies in red cell size (anisocytosis) and shape (poikilocytosis). Fifteen weeks after transplantation with unselected TNS9-transduced Hbb^{th3/+} bone marrow, the haematocrit level, red blood cell count, reticulocyte count and haemoglobin level were markedly improved in five out of five recipient mice (Fig. 6). Anisocytosis and poikilocytosis were markedly reduced in the peripheral blood smears of chimaeras bearing the TNS9 vector (Fig. 5c, d). Control mice transplanted with Hbb^{th3/+} bone marrow cells transduced with a vector encoding enhanced green fluorescent protein (eGFP) all remained severely anaemic ($n = 5$, Fig. 6) and maintained their abnormal red cell morphology (Fig. 5a, b). These results establish that levels of circulating haemoglobin obtained with TNS9 were indeed therapeutically relevant.

The combined effect of the high efficiency of gene transfer and the absence of vector rearrangements afforded by the recombinant lentivirus carrying the β -globin gene and LCR configuration adopted in TNS9 yielded levels of human β^A expression in the therapeutic range. The higher expression obtained with TNS9 compared with RNS1 was associated with a higher fraction of permissive integration sites in MEL cells (Fig. 2b) and a higher fraction of human β^A -containing red blood cells in bone marrow chimaeras (Table 1 in Supplementary Information). RNS1, which carries a weaker enhancer, silenced over time whereas TNS9 retained undiminished transcriptional activity over the same time period (Fig. 3c, d), and in secondary transplant recipients (Table 2 in Supplementary Information).

Higher levels of murine α_2 : human β_2^A tetramers were obtained in peripheral blood samples from recipients of TNS9-transduced Hbb^{th3/+} bone marrow ($21 \pm 3\%$ of total haemoglobin, $n = 5$, Fig. 3f) than with Hbb^{+/+} bone marrow ($6 \pm 4\%$, $n = 10$, Fig. 3e). The two groups showed comparable peripheral blood vector copy numbers and levels of human β -globin RNA (0.8 ± 0.2 compared with 0.8 ± 0.6 , and $16.8 \pm 6\%$ compared with $10.8 \pm 7\%$, respectively). This observation is consistent with a competitive advantage of murine β -globin over human β -globin in associating with murine α -globin²². In thalassaemic patients, added human β -

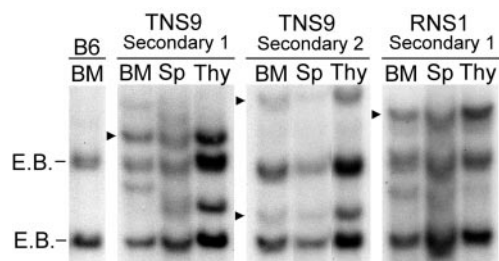


Figure 4 Integration of lentiviral vectors into hematopoietic stem cells. Southern blot analysis was performed on genomic DNA isolated from bone marrow (BM), spleen (Sp) and thymus (Thy) of secondary bone marrow transplant recipients collected 13 weeks after transplant (one representative RNS1, and two representative TNS9 secondary transplant recipients are shown). Arrowheads indicate unique integration bands found in all tissues. Two endogenous bands (E.B.) are found in the genomic DNA of C57BL/6 (B6) mice. See Table 1 in Supplementary Information for further analyses of engraftment and human β -globin expression in secondary transplant recipients.

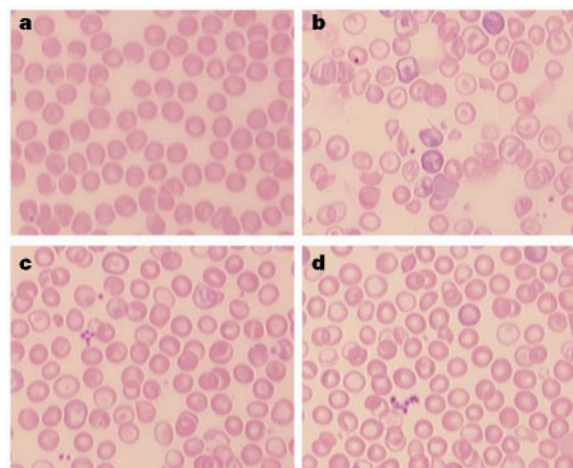


Figure 5 Correction of anisocytosis and poikilocytosis in bone marrow chimaeras reconstituted with TNS9-transduced Hbb^{th3/+} bone marrow cells. **a**, Blood smear of an adult C57BL/6 mouse transplanted with control GFP-transduced C57BL/6 bone marrow. **b**, Blood smear of an adult, Hbb^{th3/+} mouse transplanted with GFP-transduced Hbb^{th3/+} bone marrow. **c, d**, Blood smear of two TNS9-transduced Hbb^{th3/+} bone marrow chimaeras. Blood smears were taken 7 weeks after transplantation with transduced bone marrow.

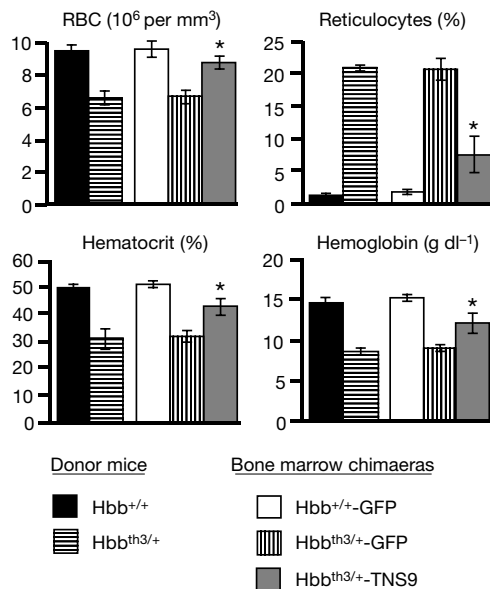


Figure 6 Amelioration of haematological parameters in bone marrow chimaeras reconstituted with TNS9-transduced Hbb^{th3/+} bone marrow cells. Red blood cell (RBC) counts, reticulocyte counts, hematocrit and hemoglobin levels are shown 15 weeks after transplantation ($n = 5$ mice per group). All measured parameters show significant increases (asterisk) in TNS9-transduced Hbb^{th3/+} as compared with GFP-transduced Hbb^{th3/+} bone marrow chimaeras ($P = 0.1, 0.2, 0.1$, and 0.4 , respectively; Wilcoxon rank sum test).

chain synthesis would improve the α : β chain imbalance and thus increase red cell survival and ameliorate the ineffective erythropoiesis in these patients. In patients with sickle cell disease, transduced β^A chains are expected to have an advantage over the β^S chains produced by both endogenous genes in competing for the available α -chains²³. Given that patients with S/ β -thalassaemia whose HbA represents 10–30% of their total haemoglobin are very mildly affected^{1,24}, the clinical benefit of such an intervention would be highly significant.

We have shown that recombinant lentiviruses bearing large fragments of the human β -globin gene and its LCR represent an advancement towards the genetic treatment of severe haemoglobinopathies. Lentiviral vectors appear to be well suited for the stable transduction of human CD34⁺ cells²⁵ and are expected to be available for therapeutic applications once safety concerns are fully addressed²⁶. Furthermore, the principles underlying inclusion of multiple genetic elements within this vector provide a paradigm for any stem cell therapy requiring stable and regulated expression of a tissue-specific transgene. □

Methods

Vector construction and production

The human β -globin gene was subcloned from M β 6L (ref. 9) into pHR' LacZ (ref. 27; kindly provided by D. Trono) replacing the CMV–LacZ sequence. pHR'eGFP was constructed by replacing LacZ with the eGFP coding sequence (Clontech). Viral stocks were generated by triple transfection of the recombinant vectors, pCMV Δ R8.9 (ref. 27) and pMD.G into 293T cells as described²⁸. The pseudotyped virions were concentrated by ultracentrifugation, resuspended and titrated as described²⁹.

Northern and Southern blot analyses

We transiently transfected 293T cells with 15 μ g of the packaging vector pCMV Δ R8.9 and 20 μ g of RNS1 or TNS9 plasmid. Total RNA was isolated 48 h after transfection with TRIzol (Gibco BRL). Northern blot analyses were done as described³⁰ with the NcoI–BamHI fragment of the human β -globin gene as probe. Cell-free viral supernatant was used to infect HeLa cells in the presence of polybrene (8 μ g ml⁻¹). For analysis of integrated vector structure, genomic DNA from cells was isolated, digested with ScaI, and studied by Southern blot analysis⁹ using a [³²P]dCTP-labelled NcoI–BamHI fragment of

the human β -globin gene as probe. Plasmid controls were also digested with ScaI and admixed with 5 μ g of ScaI-digested HeLa genomic DNA. Quantification of vector copy number in bone marrow chimaeras was performed on genomic DNA digested with BamHI. Analysis of vector integration sites in secondary bone marrow transplant recipients was performed on samples digested with NcoI, which cuts only once in the vector.

MEL cell infection, subcloning and differentiation

Cell-free viral supernatant was used to infect C8H MEL cells⁹ in the presence of polybrene (8 μ g ml⁻¹). Transduced MEL cells were subcloned by limiting dilution, and screened by PCR for transduction³⁰ using primers that anneal in the human β -globin promoter sequence (BPS, 5'-GTCTAAGTGATGACAGCCGTACCTG-3') and in HS2 (C2A, 5'-TCAGCCTAGAGT GATGACTCC TATCTG-3'). Vector copy number and integration site analysis was determined by Southern blot analysis⁹. Transduced MEL cells were induced to maturation by 5-day culture in 5 mM N,N'-hexamethylene bisacetamide (HMBA, Sigma).

Quantification of human β -globin mRNA

Total RNA was extracted from MEL, Jurkat and HL-60 cells, or mouse spleen and blood using TRIzol. Quantitative primer extension assays were done using the Primer Extension System-AMV Reverse Transcriptase kit (Promega) with [³²P]dATP end-labelled primers specific for retroviral-derived human β -globin (5'-CAGTAACGGCAGACTTCTCTCTC-3') and mouse β -globin (5'-TGATGTCTGTTTCTGGGGTT GTG-3'), with predicted extension products of 90 bp and 53 bp, respectively. The probes yield products of identical length for β^{maj} , β^{min} , β^S and β^A . Primers were annealed to 4 μ g of RNA and reactions were run according to manufacturer's protocols. Radioactive bands were quantitated by phosphorimager analysis (BioRad). RNA isolated from A85.68 mice³⁰ was used as positive control. After correction for primer labelling, the human to mouse RNA signal was $29 \pm 1\%$ per gene copy in repeated experiments ($n > 8$), in agreement with previous findings based on RT-PCR²⁰. Values measured in bone marrow chimaeras that were obtained in separate runs were standardized to the value obtained in the A85.68 RNA sample. In Figs 2 and 3c, d, human β -globin expression is expressed per vector copy and normalized to the endogenous transcripts (accounting for two endogenous alleles). In Fig. 3b, human transcripts are reported as the fraction of total β -globin RNA (Hu β / Hu β + Mu β) to reflect absolute contribution of vector-encoded transcripts.

Bone marrow chimaeras

Donor bone marrow was flushed from the femurs of 8- to 16-week-old male C57BL/6 or Hbb^{th3/+} mice (Jackson Laboratories) that had been injected intravenously (i.v.) 6 days earlier with 5-fluorouracil (5-FU, Pharmacia; 150 mg kg⁻¹ body weight). Bone marrow cells were resuspended in serum-free medium, and supplemented with IL-1 α (10 ng ml⁻¹), IL-3 (100 U ml⁻¹), IL-6 (150 U ml⁻¹), Kit ligand (10 ng ml⁻¹) (Genzyme), β -mercaptoethanol (0.5 mM; Sigma), L-glutamine (200 mM), penicillin (100 IU ml⁻¹) and streptomycin (100 μ g ml⁻¹), and cultured for 18 h. Recipient mice (11- to 14-week-old C57BL/6 or Hbb^{th3/+} mice) were irradiated with 10.5 Gy (split dose 2×5.25 Gy) on the day of transplantation. Bone marrow cells were pelleted and resuspended in serum-free medium containing concentrated lentiviral supernatant, and supplemented with polybrene (8 μ g ml⁻¹), L-glutamine (200 mM), penicillin (100 IU ml⁻¹) and streptomycin (100 μ g ml⁻¹), and cultured for 6 h. Transduced bone marrow cells (1×10^5 or 5×10^5) were then i.v. injected into each of the irradiated female recipients to establish short-term and long-term bone marrow chimaeras, respectively.

In short-term studies, spleens were removed 12 d after transplantation to extract total RNA and genomic DNA. To monitor long-term chimaeras, two or three capillary tubes of blood were collected every 4–6 weeks, from which genomic DNA, total RNA and haemoglobin were extracted. To examine vector function reliably in long-term animals, erythroid cell populations were purified from spleen. Single-cell suspensions were incubated with rat anti-mouse TER-119 monoclonal antibody (PharMingen). Sheep anti-Rat IgG dynabeads (M-450, Dynal Inc.) were added to the antibody-coated spleen cells and purified as recommended by the manufacturer. Vector copy number, integration pattern and chimaerism were determined by Southern blot analysis. The fraction of donor DNA relative to recipient was determined by stripping and reprobing the blot with a [³²P]dCTP-labelled probe specific for the Y chromosome and normalizing to an endogenous mouse band. Radioactive bands were quantitated by phosphorimager analysis. Sera from five randomly selected long-term bone marrow chimaeras (30 weeks after transplantation) tested negative for HIV-1 gag by RT-PCR using the Amplicor HIV-1 monitor kit (Roche).

Protein analyses

Red cell lysates of freshly collected peripheral blood were analysed by cellulose acetate electrophoresis (pH 8.5, Helena Laboratories). The fraction of Hbb^{hu} relative to total Hbb (Hbb^{hu} / Hbb^{hu} + Hbb^{mu}) was calculated by running four serial dilutions and measuring each band by densitometry in the linear range. To measure the fraction of peripheral blood cells expressing human β^A , smears were stained with PE-conjugated anti-mouse TER-119 α monoclonal antibody and FITC-conjugated monoclonal antibody to human haemoglobin A (EG&G WALLAC) as recommended by the manufacturer. The slides were analysed with an immunofluorescence microscope (Olympus BX60).

Red cell morphology and haematologic studies

Blood smears were stained with Wright-Giemsa. Total haemoglobin, red cell counts and

haematocrits were measured on a CBC analyser (H System, Bayer Corporation). Reticulocytes were counted after staining with New Methylene Blue, Brecher Formula (JT Baker).

Statistics

We used the non-parametric Wilcoxon sum rank test procedure to determine whether human β -globin expression or haematological parameters differed between different treatment groups. Fisher's exact test was used to determine whether two proportions in the population were equal. A low value of P is evidence that the two proportions are different.

Received 10 February; accepted 17 May 2000.

- Weatherall, D. J. & Clegg, J. B. *The Thalassemia Syndromes* (Blackwell Scientific, Oxford, 1981).
- Nathan, D. G. & Oski, F. A. *Hematology of Infancy and Childhood* (WB Saunders, Philadelphia, 1998).
- Sadelain, M. Genetic treatment of the haemoglobinopathies: recombinations and new combinations. *Br. J. Haematol.* **98**, 247–253 (1997).
- Boulad, F. *et al.* Bone marrow transplantation for homozygous beta-thalassemia. *Ann. N. Y. Acad. Sci.* **850**, 498–502 (1998).
- Rivella, S. & Sadelain, M. Genetic treatment of severe hemoglobinopathies: the combat against transgene variegation and transgene silencing. *Semin. Hematol.* **35**, 112–125 (1998).
- Tuan, D., Solomon, W., Li, Q. & London, I. M. The "beta-like-globin" gene domain in human erythroid cells. *Proc. Natl Acad. Sci. USA* **82**, 6384–6388 (1985).
- Forrester, W. C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. & Groudine, M. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucleic Acids Res.* **15**, 10159–10177 (1987).
- Grosfeld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**, 975–985 (1987).
- Sadelain, M., Wang, C. H., Antoniou, M., Grosfeld, F. & Mulligan, R. C. Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. *Proc. Natl Acad. Sci. USA* **92**, 6728–6732 (1995).
- Leboulch, P. *et al.* Mutagenesis of retroviral vectors transducing human beta-globin gene and beta-globin locus control region derivatives results in stable transmission of an active transcriptional structure. *EMBO J.* **13**, 3065–3076 (1994).
- Raftopoulos, H., Ward, M., Leboulch, P. & Bank, A. Long-term transfer and expression of the human beta-globin gene in a mouse transplant model. *Blood* **90**, 3414–3422 (1997).
- Grosfeld, F. *et al.* The dynamics of globin gene expression and gene therapy vectors. *Semin. Hematol.* **35**, 105–111 (1998).
- Bulger, M. & Groudine, M. Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* **13**, 2465–2477 (1999).
- Higgs, D. R. Do LCRs open chromatin domains? *Cell* **95**, 299–302 (1998).
- Engel, J. D. & Tanimoto, K. Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. *Cell* **100**, 499–502 (2000).
- Novak, U., Harris, E. A., Forrester, W., Groudine, M. & Gelinas, R. High-level beta-globin expression after retroviral transfer of locus activation region-containing human beta-globin gene derivatives into murine erythroleukemia cells. *Proc. Natl Acad. Sci. USA* **87**, 3386–3390 (1990).
- Talbot, D. *et al.* A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature* **338**, 352–355 (1989).
- Bungert, J. *et al.* Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev.* **9**, 3083–3096 (1995).
- Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V. & Cullen, B. R. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338**, 254–257 (1989).
- Porcu, S. *et al.* The human beta globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. *Blood* **90**, 4602–4609 (1997).
- Yang, B. *et al.* A mouse model for beta 0-thalassemia. *Proc. Natl Acad. Sci. USA* **92**, 11608–11612 (1995).
- Fabry, M. E. *et al.* A second generation transgenic mouse model expressing both hemoglobin S (HbS) and HbS-Antilles results in increased phenotypic severity. *Blood* **86**, 2419–2428 (1995).
- Esan, G. J., Adesina, T. A. & Luzzatto, L. Synthesis of haemoglobins specified by allelic genes in human heterozygotes. *Nature* **229**, 143–145 (1971).
- Stamatoyannopoulos, G., Nienhuis, A. W., Majerus, P. & Varmus, H. *The Molecular Basis Of Blood Diseases* (WB Saunders, Philadelphia, 1994).
- Miyoshi, H., Smith, K. A., Mosier, D. E., Verma, I. M. & Torbett, B. E. Transduction of human CD34+ cells that mediate long-term engraftment of NOD/SCID mice by HIV vectors. *Science* **283**, 682–686 (1999).
- Amado, R. G. & Chen, I. S. Lentiviral vectors—the promise of gene therapy within reach? *Science* **285**, 674–676 (1999).
- Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nature Biotechnol.* **15**, 871–875 (1997).
- Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**, 8463–8471 (1998).
- Gallardo, H. F., Tan, C., Ory, D. & Sadelain, M. Recombinant retroviruses pseudotyped with the vesicular stomatitis virus G glycoprotein mediate both stable gene transfer and pseudotransduction in human peripheral blood lymphocytes. *Blood* **90**, 952–957 (1997).
- Rivella, S., Callegari, J. A., May, C., Tan, C. & Sadelain, M. The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. *J. Virol.* **74**, 4679–4687 (2000).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank M. Trudel for helpful discussion; C. Tan and H. Beauchemin for technical

assistance and I. Rivière for reviewing the manuscript. This work was supported by grants from the NHLBI, the NCI, the Cancer Research Institute, the Cooley's Anaemia Foundation, DeWitt-Wallace Fund and the McDonnell Foundation Scholars Award.

Correspondence and requests should be addressed to M.S. (e-mail: m-sadelain@ski.mskcc.org).

Requirement for glycogen synthase kinase-3 β in cell survival and NF- κ B activation

Klaus P. Hoeflich*, Juan Luo*, Elizabeth A. Rubie, Ming-Sound Tsao, Ou Jin & James R. Woodgett

Ontario Cancer Institute/Princess Margaret Hospital, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada

* These authors contributed equally to this work

Glycogen synthase kinase-3 (GSK-3)- α and - β are closely related protein-serine kinases, which act as inhibitory components of Wnt signalling during embryonic development and cell proliferation in adult tissues^{1,2}. Insight into the physiological function of GSK-3 has emerged from genetic analysis in *Drosophila*^{3,4}, *Dictyostelium*⁵ and yeast^{6,7}. Here we show that disruption of the murine GSK-3 β gene results in embryonic lethality caused by severe liver degeneration during mid-gestation, a phenotype consistent with excessive tumour necrosis factor (TNF) toxicity, as observed in mice lacking genes involved in the activation of the transcription factor activation NF- κ B. GSK-3 β -deficient embryos were rescued by inhibition of TNF using an anti-TNF- α antibody. Fibroblasts from GSK-3 β -deficient embryos were hypersensitive to TNF- α and showed reduced NF- κ B function. Lithium treatment (which inhibits GSK-3; refs 8, 9) sensitized wild-type fibroblasts to TNF and inhibited transactivation of NF- κ B. The early steps leading to NF- κ B activation (degradation of I- κ B and translocation of NF- κ B to the nucleus) were unaffected by the loss of GSK-3 β , indicating that NF- κ B is regulated by GSK-3 β at the level of the transcriptional complex. Thus, GSK-3 β facilitates NF- κ B function.

In vertebrates, GSK-3 is crucial for the definition of the embryonic axes^{10–12}. To investigate the role of GSK-3 in mammalian development, we disrupted the GSK-3 β gene in murine 129J embryonic stem (ES) cells using a targeting vector in which the exon encoding the ATP-binding loop was deleted (Fig. 1a). Chimaeric mice derived from two independent heterozygous ES clones were back-crossed to C57BL/6J mice, and heterozygous mice were crossed to generate homozygous mutant offspring. The null mutation of GSK-3 β was confirmed by tail DNA genotyping and Southern (Fig. 1b) and western blot (Fig. 1c) analyses of embryonic fibroblasts derived from mice on day 12.5 of gestation (E12.5). Although GSK-3 β ^{+/-} male and female mice were healthy and fertile, they did not give rise to live GSK-3 β ^{-/-} progeny.

We analysed embryos from timed pregnancies of GSK-3 β ^{+/-} intercrosses. Before E12.5, the mendelian ratio of null embryos was normal and most GSK-3 β ^{-/-} embryos showed no morphological abnormality (not shown). Between E13.5 and E14.5, however, GSK-3 β ^{-/-} embryos were pale and non-viable (Fig. 2a). Histological examination of E13.5 GSK-3 β ^{+/-} and GSK-3 β ^{-/-} embryos showed multifocal haemorrhagic degeneration in the livers of homozygous embryos (Fig. 2b). In these areas, the hepatocytes were pyknotic and karyorrhectic, indicating that they were probably undergoing

apoptosis. This was confirmed by terminal deoxynucleotidyl transferase-mediated deoxyuridine triphosphate nick-end labelling (TUNEL) assay, which showed many positively stained nuclei (Fig. 2b). *GSK-3 β* ^{-/-} embryos derived from the two original *GSK-3 β* heterozygous ES cell clones were indistinguishable in phenotype (not shown). Foetal liver transfer experiments showed that haematopoietic cells from *GSK-3 β* ^{-/-} embryos were capable of reconstituting haematopoiesis in lethally irradiated mice (not shown). There was no evidence that Wnt signalling, cyclin D levels, β -catenin accumulation or glycogen metabolism were perturbed (not shown). These results indicate that hepatocyte apoptosis is probably the major cause of lethality in the *GSK-3 β* ^{-/-} embryos.

This phenotype is similar to that of IKK β or RelA-deficient mice, which results from increased sensitivity to TNF- α ¹³⁻¹⁵. To inhibit endogenous TNF- α , we injected an anti-murine TNF- α monoclonal hamster antibody (or hamster immunoglobulins as a control) into the peritoneum of three pregnant heterozygous females (sired by heterozygous males) on E10.5 or E11.5. The females were killed and the embryos collected on E14.5. Eleven embryos (42% of total) were identified as *GSK-3 β* ^{-/-}. The gross morphology (not shown) and histology of the liver showed that the anti-TNF- α injections produced substantial protection and hepatocyte rescue (Fig. 2c). These results pointed to an unexpected role for *GSK-3 β* in suppressing TNF-mediated apoptosis.

To investigate how *GSK-3 β* affected TNF cytotoxicity, we examined immortalized mouse embryonic fibroblast cell lines from two wild-type (+/+1 and +/+2), two *GSK-3 β* heterozygous (+/-1 and +/-2) and two *GSK-3 β* homozygous (-/-1 and -/-2) mutant embryos. The expression of proteins important for TNF signalling, such as TNF-R1 and IKK β , was normal in these lines (not shown). Following TNF- α treatment (10 ng ml⁻¹, 24 h), we visualized the nuclear morphology of *GSK-3 β* ^{+/+} and *GSK-3 β* ^{-/-} cells by acridine orange staining and fluorescence microscopy. The chromatin condensation characteristic of apoptosis was apparent only in the *GSK-3 β* ^{-/-} cell lines (Fig. 3a). We then tested the susceptibility of wild-type and *GSK-3 β* -deficient cells to TNF in the absence or presence of cycloheximide (CHX; Fig. 3b). As

measured by viability dye staining, wild-type and *GSK-3 β* -heterozygous cells were relatively unaffected by exposure to TNF- α . However, significantly fewer *GSK-3 β* ^{-/-} cells were viable 24 h after maximum treatment with TNF (18% and 56%) or with TNF/CHX (2% and 28%, respectively). As a reference, TNF receptor-associated

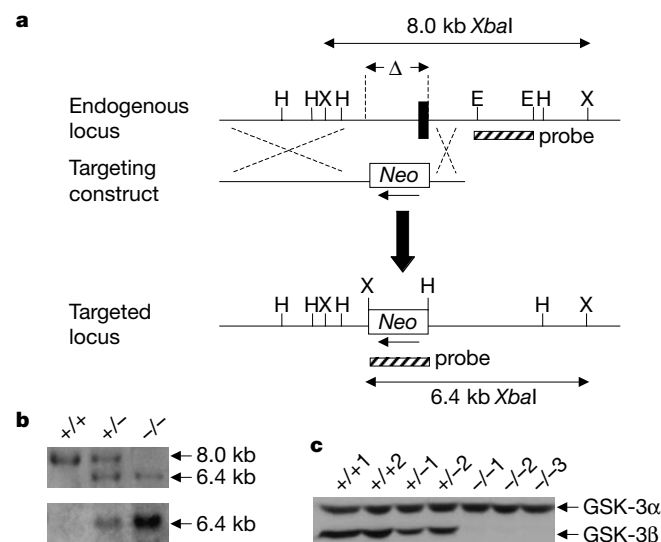


Figure 1 Targeted disruption of the *GSK-3 β* locus. **a**, A portion of the mouse *GSK-3 β* wild-type locus (top) showing exons (solid box) and a 8.0-kilobase (kb) *XbaI* fragment in the wild-type allele. The mutated *GSK-3 β* locus (bottom) contains a 6.4-kb *XbaI* fragment. The positions of the probes used for Southern analysis are shown (hatched boxes). X, E and H represent *XbaI*, *EcoRI* and *HindIII* restriction sites, respectively. **b**, Southern blot of embryonic fibroblast DNA digested with *XbaI* and hybridized to the flanking probe. **c**, *GSK-3 β* protein is undetectable in *GSK-3 β* ^{-/-} embryonic fibroblasts.

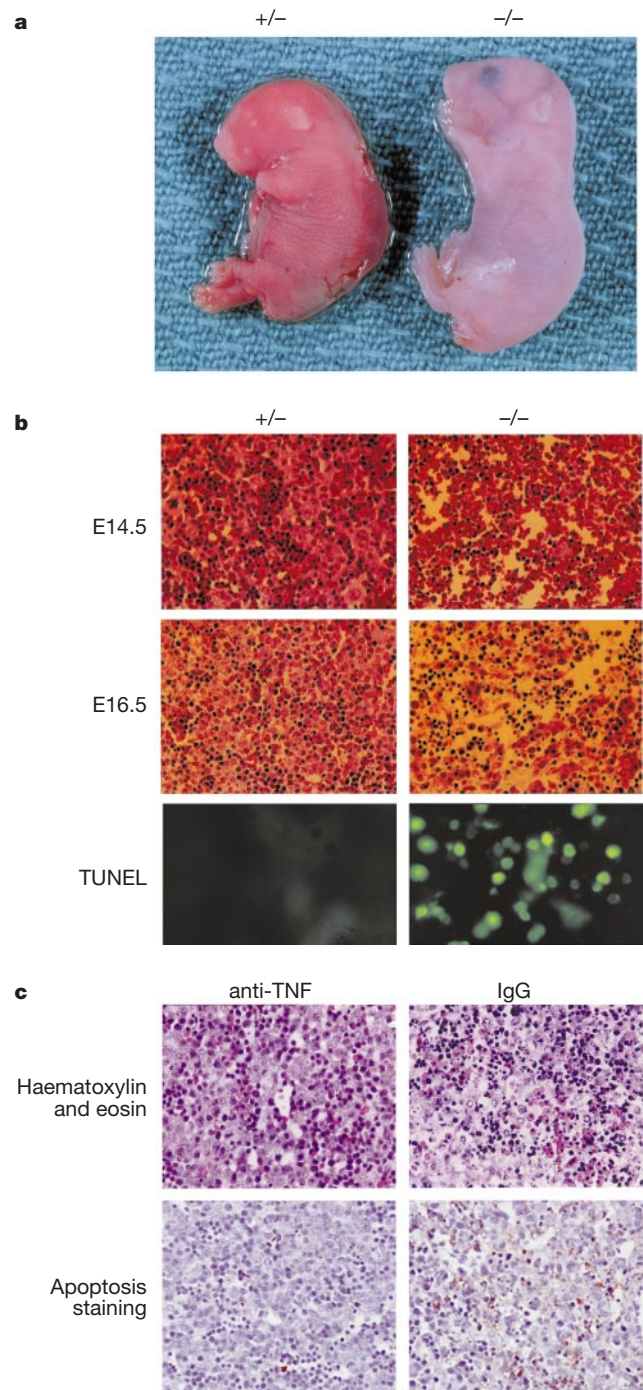


Figure 2 Phenotype of *GSK-3 β* ^{-/-} embryos. **a**, *GSK-3 β* ^{-/-} embryos at E14.5 showed focal parenchymal haemorrhage. **b**, Histological sections of *GSK-3 β* ^{-/-} embryos stained with haematoxylin and eosin showed liver abnormality with nuclear debris and apoptotic cells at varying stages of degeneration. Increased TUNEL-positive cells were observed in *GSK-3 β* ^{-/-} mice. **c**, Pregnant females from *GSK-3 β* ^{+/+} × *GSK-3 β* ^{+/+} intercrosses were injected intraperitoneally with 250 μ g anti-murine TNF- α hamster monoclonal antibody or with control hamster IgG. Liver section analysis was performed on *GSK-3 β* -null genotyped embryos and apoptosis was detected by *in situ* end-labelling of fragmented DNA.

factor 2 (TRAF2) homozygous mutant mouse embryonic fibroblasts were used in the same assay. *TRAF2*^{-/-} cell survival was consistent with previous report¹⁶ (~20% at 10 ng ml⁻¹ TNF- α /0.25 μ g ml⁻¹ CHX).

Lithium ions inhibit GSK-3 α and - β both *in vitro* and in intact cells⁹ and have been reported to exacerbate TNF cytotoxicity¹⁷. We incubated cells with 20 mM lithium (or with 20 mM potassium as control) during treatment with CHX and an intermediate TNF- α concentration (10 ng ml⁻¹). Consistent with the hypersensitivity of the GSK-3 β null cells to TNF- α , treatment of wild-type and heterozygous cell lines with the combination of lithium/TNF- α /

CHX resulted in a similar increase in TNF cytotoxicity (Fig. 3c).

To demonstrate that the apoptotic sensitivity of *GSK-3 β* ^{-/-} fibroblasts was a direct consequence of GSK-3 β deficiency, a rat *GSK-3 β* gene was introduced into *GSK-3 β* ^{-/-} cells by transient transfection. *GSK-3 β* ^{-/-} fibroblasts expressing the *GSK-3 β* transgene were less sensitive to TNF- α treatment than controls (Fig. 4a). We assessed the specificity of apoptotic sensitivity upon disruption of GSK-3 β function by treating *GSK-3 β* ^{-/-} cells with a range of pro-apoptotic agents (anisomycin, sorbitol, daunorubicin or γ -irradiation); no differences from wild-type responses were observed (Fig. 4b). These data indicate that GSK-3 β is required for protection from TNF- α cytotoxicity.

Activation of the transcription factor NF- κ B is essential for the suppression of TNF-induced apoptosis^{18–21}. To determine whether TNF-induced NF- κ B activation was affected in the *GSK-3 β* ^{-/-} cells, nuclear extracts from TNF- α -treated embryonic fibroblast cell lines were analysed by electrophoretic mobility shift assay (EMSA) using an NF- κ B-specific probe. The specificity of the shifted band was tested using a 200-fold excess of non-labelled wild-type or mutant κ B DNA (Fig. 5a). NF- κ B activation in response to TNF- α treatment was reduced by more than 50% in *GSK-3 β* ^{-/-} cells, as

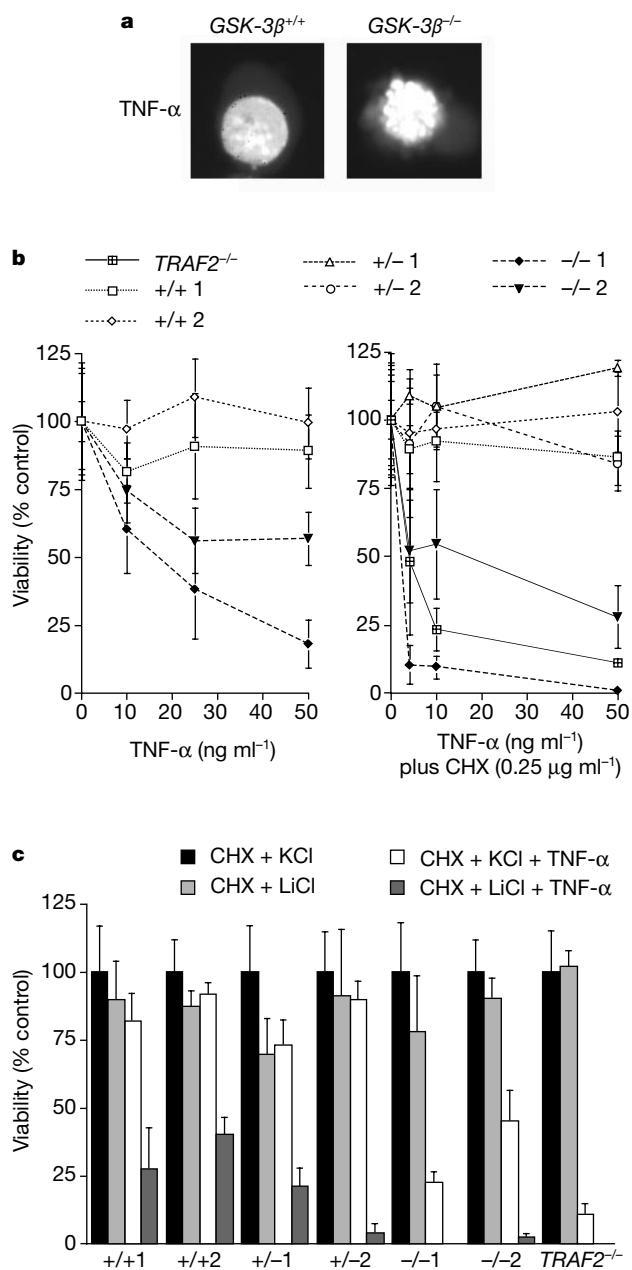


Figure 3 *GSK-3 β* -deficient cells show increased sensitivity to TNF-induced apoptosis. **a**, Chromatin condensation demonstrated by acridine orange staining. **b**, Cell viability following 24 h TNF- α (10, 25 or 50 ng ml⁻¹) treatment. *GSK-3 β* ^{+/+} embryonic fibroblast cell clones are indicated by $\pm$ 1 and $\pm$ 2; *GSK-3 β* ^{-/-} by $-$ 1 and $-$ 2. Survival after treatment with TNF- α (4, 10 or 50 ng ml⁻¹) plus CHX (0.25 μ g ml⁻¹) is shown in the right panel. **c**, Lithium potentiates TNF-induced death in wild-type cells. Cells were treated with 0.25 μ g ml⁻¹ CHX, 20 mM K⁺, 20 mM Li⁺ and 10 ng ml⁻¹ TNF- α for 24 h as indicated.

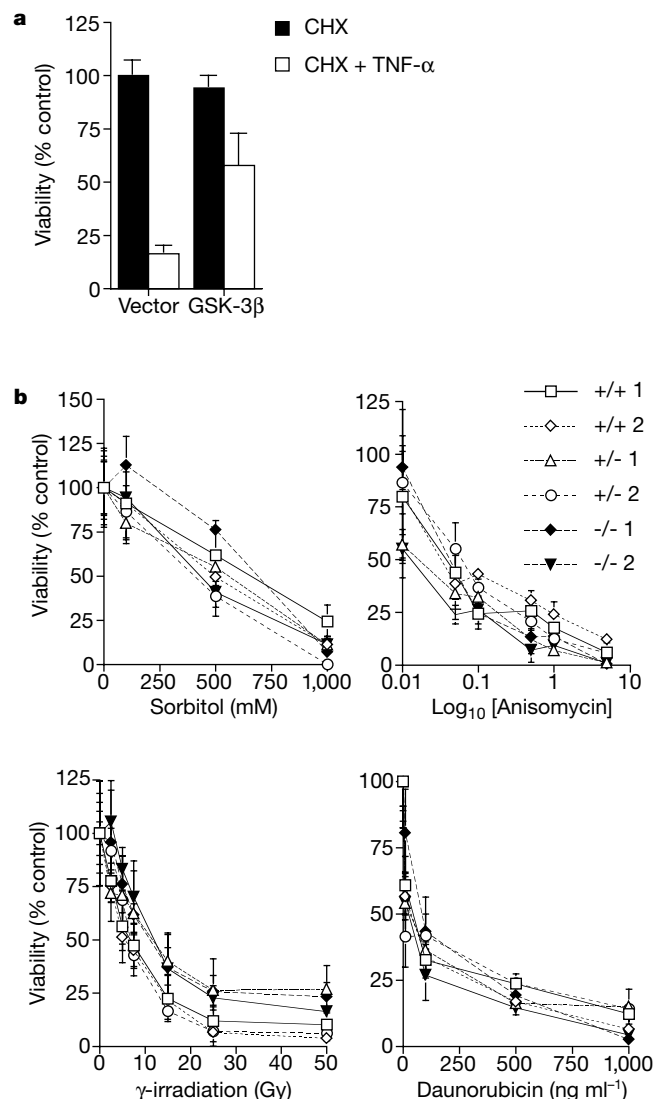


Figure 4 Apoptotic sensitivity of *GSK-3 β* ^{-/-} fibroblasts. **a**, Expression of exogenous *GSK-3 β* in *GSK-3 β* ^{-/-} cells protects against TNF-induced apoptosis. **b**, Sensitivity of *GSK-3 β* ^{-/-} embryonic fibroblast cells to various apoptotic stimuli.

compared to wild-type cells (Fig. 5b). The effect of GSK-3 β disruption was specific to NF- κ B binding, as the DNA binding of other transcription factors such as Oct-1 was unaffected (Fig. 5c).

NF- κ B-mediated transcription of the endothelial leukocyte adhesion molecule-1 (ELAM-1 or E-selectin) gene is induced by TNF- α and interleukin-1 β (IL-1 β). We therefore examined ELAM-luciferase induction by these agonists in embryonic fibroblast cells. Induction of ELAM-luciferase transcription by TNF or IL-1 β was reduced approximately 3–4-fold in GSK-3 β ^{-/-} cells (Fig. 5d, e) and could be restored by the expression of exogenous GSK-3 β (Fig. 5f). The diminished induction of NF- κ B by IL-1 β in GSK-3 β ^{-/-} cells indicates that the defect caused by lack of GSK-3 β is in NF- κ B function, rather than in a TNF-specific response. Lithium inhibited the induction of NF- κ B transactivation by TNF- α by about 70% in HEK293 epithelial cells (Fig. 5g).

In resting cells, NF- κ B complexes are retained in the cytoplasm by association with a family of inhibitory proteins termed I- κ Bs²². Activation of NF- κ B requires the phosphorylation of I- κ B, which triggers its polyubiquitination and subsequent degradation by the 26S proteasome. The liberated NF- κ B rapidly translocates to the nucleus, where it binds κ B sites and regulates gene expression. To

assess how NF- κ B activity was inhibited in GSK-3 β null cells, we studied the kinetics of I- κ B proteolysis and NF- κ B translocation following treatment of cells with TNF- α . Western blotting of cytoplasmic and nuclear extracts indicated no differences between wild-type and GSK-3 β -deficient cells (Fig. 5h). In all cells lines tested, I- κ B- α was rapidly degraded following exposure to TNF- α . This was followed by a lag phase and reaccumulation of the protein 2 h later. Translocation of the p65 subunit of NF- κ B to the nucleus was also independent of GSK-3 β . These findings place the defect in NF- κ B activation downstream of I- κ B phosphorylation and the nuclear translocation of NF- κ B. In addition, they imply that GSK-3 β activity is required for nuclear activity of the complex. As I- κ B synthesis—which is under the transcriptional control of NF- κ B—is unchanged in GSK-3 β ^{-/-} cells, only a subset of NF- κ B-inducible genes are dependent on GSK-3 β function. As described previously, partial inhibition of NF- κ B DNA binding alone does not seem to be sufficient for downregulation of the I- κ B- α gene²³. Activation of phosphatidylinositol-3-OH kinase protein kinase B/Akt can result in activation of NF- κ B^{24–28} as well as in partial inactivation of GSK-3 (ref. 29). The relationships between these opposing effects (analogous to the I- κ B cycle) require further investigation.

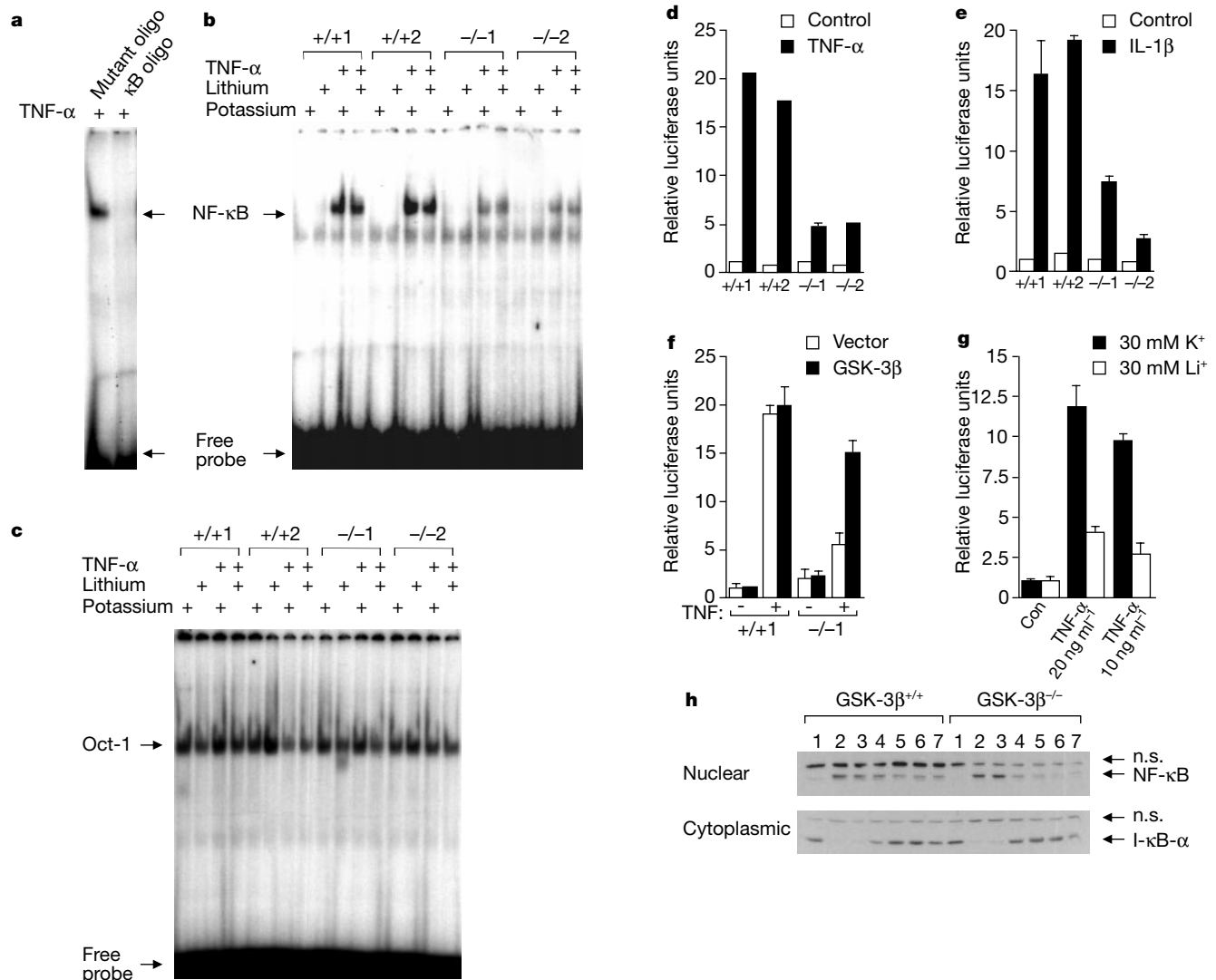


Figure 5 Inhibition of GSK-3 β reduces TNF-inducible NF- κ B activation. **a**, Specificity was evaluated by oligonucleotide competition and EMSA. **b**, Activation of NF- κ B in GSK-3 β ^{+/+} and GSK-3 β ^{-/-} cells treated with TNF- α . **c**, GSK-3 β does not interfere with DNA binding of Oct-1. **d**, **e**, Effect of GSK-3 β disruption on TNF- α (**d**) and IL-1 β signalling (**e**).

f, **g**, GSK-3 β deficiency (**f**) or lithium treatment (**g**) block NF- κ B-mediated gene transcription. **h**, Immunoblot analysis of cytoplasmic I- κ B- α and nuclear p65 NF- κ B. Cells were stimulated with 10 ng ml⁻¹ TNF- α for 0, 0.25, 0.5, 1, 2, 4 or 8 h (lanes 1–7). Non-specific (n.s.) bands are indicated.

In summary, we have shown that GSK-3 β function is required for the NF- κ B-mediated anti-apoptotic response to TNF- α . Our data also show that GSK-3 α and - β have distinct biological roles, as the former is unable to compensate for the loss of the latter. □

Methods

Cytoplasmic and nuclear lysates were prepared as described¹⁶. Immunoblotting was carried out with I- κ B- α (polyclonal, New England Biolabs), p65 NF- κ B (polyclonal, Santa Cruz Biotechnology) and GSK-3 (mouse monoclonal, Upstate Biotechnology) antibodies.

Apoptosis assays

Where indicated, TNF- α -treated cells (10 ng ml⁻¹ h) were collected, mixed with 4 μ g ml⁻¹ acridine orange (final concentration) and assessed by fluorescence microscopy. Cell survival following TNF- α treatment was determined by negative staining with trypan blue and expressed normalized to untreated controls. All experiments were repeated at least three times, and the data are shown as the mean $\pm$ standard error. *In situ* apoptosis was detected by TUNEL assay according to the manufacturer's instructions (Boehringer Mannheim), or by another fragmented DNA end-labelling protocol³⁰. For β -galactosidase viability assays, two days after transfection with pCMV- β -galactosidase plus either pCDNA3-HA-GSK-3 β or control plasmid, cells were treated as indicated and analysed.

EMSA

The κ B-binding activities of embryonic fibroblasts incubated with lithium or potassium (30 mM, 4 h) and murine TNF- α (100 ng ml⁻¹, 30 min), as indicated, were compared by EMSA. Nuclear lysates were prepared and EMSAs were performed as described¹⁷. For oligonucleotide competition assays, equivalent amounts of nuclear extract protein (3 μ g) were preincubated for 5 min with a 200-fold excess of either NF- κ B-specific oligonucleotide probe containing two tandem NF- κ B-binding sites (5'-ATCAGGGACTTTCCCGTGGGGACTTTCCG-3' and 5'-CGGAAAGTCCCCAGCGAAAGTCCCTGAT-3') or mutant NF- κ B oligonucleotides (5'-GATCACTCACTTTCCGCTTGCTCACTTTCCAG-3' and 5'-CTGGAAGTGAGCAAGCGCAAAGTGAGTGATC-3') before addition of the radiolabelled NF- κ B using the oligonucleotides 5'-TTCTAGTGATTGCATTCGACA-3' and 5'-TGTCGAATGCAATCACTAGAA-3'.

Luciferase assay

Embryonic fibroblast cells were transfected with plasmids expressing ELAM-luciferase and β -galactosidase. Transfected cells were incubated in the presence of 30 ng ml⁻¹ TNF- α or IL-1 β for 6 h. Luciferase assays were carried out using the Promega assay kit and a Berthold luminometer. Activity was normalized to β -galactosidase activity and plotted as the mean $\pm$ standard deviation of triplicates from a representative experiment. To examine the effect of lithium treatment on NF- κ B-mediated gene transcription, HEK293 epithelial cells were preincubated overnight with 30 mM lithium or potassium before being stimulated with 10 or 20 ng ml⁻¹ TNF- α . Luciferase values were normalized to the unstimulated potassium and lithium controls.

Received 31 January; accepted 19 April 2000.

- Welsh, G. I., Wilson, C. & Proud, C. G. GSK-3: a SHAGGY frog story. *Trends Cell Biol.* **6**, 274–279 (1996).
- Dale, T. C. Signal transduction by the Wnt family of ligands. *Biochem. J.* **239**, 209–223 (1998).
- Siegrfried, E., Perkins, L. A., Capaci, T. M. & Perrimon, N. Putative protein kinase product of the *Drosophila* segment-polarity gene *zeste-white3*. *Nature* **345**, 825–829 (1990).
- Ruel, L., Bourouis, M., Heitzler, P., Pantescio, V. & Simpson, P. *Drosophila* shaggy kinase and rat glycogen synthase kinase-3 have conserved activities and act downstream of Notch. *Nature* **362**, 557–560 (1993).
- Harwood, A. J., Plyte, S. E., Woodgett, J., Strutt, H. & Kay, R. R. Glycogen synthase kinase 3 regulates cell fate in *Dictyostelium*. *Cell* **80**, 139–148 (1995).
- Puziss, J. W., Hardy, T. A., Johnson, R. B., Roach, P. J. & Hieter, P. MDS1, a dosage suppressor of an mck1 mutant, encodes a putative yeast homolog of glycogen synthase kinase 3. *Mol. Cell. Biol.* **14**, 831–839 (1994).
- Plyte, S. E., Feoktistova, A., Burke, J. D., Woodgett, J. R. & Gould, K. L. *Schizosaccharomyces pombe* skp1+ encodes a protein kinase related to mammalian glycogen synthase kinase 3 and complements a cdc14 cytokinesis mutant. *Mol. Cell. Biol.* **16**, 179–191 (1996).
- Klein, P. S. & Melton, D. A. A molecular mechanism for the effect of lithium on development. *Proc. Natl Acad. Sci. USA* **93**, 8455–8459 (1996).
- Stambolic, V., Ruel, L. & Woodgett, J. R. Lithium inhibits glycogen synthase kinase-3 activity and mimics wingless signalling in intact cells. *Curr. Biol.* **6**, 1664–1668 (1996).
- He, X., Saint-Jeannet, J. P., Woodgett, J. R., Varmus, H. E. & Dawid, I. B. Glycogen synthase kinase-3 and dorsoventral patterning in *Xenopus* embryos. *Nature* **374**, 617–622 (1995).
- Dominguez, I., Itoh, K. & Sokol, S. Y. Role of glycogen synthase kinase 3 beta as a negative regulator of dorsoventral axis formation in *Xenopus* embryos. *Proc. Natl Acad. Sci. USA* **92**, 8498–8502 (1995).
- Nasevicius, A. *et al.* Evidence for a frizzled-mediated wnt pathway required for zebrafish dorsal mesoderm formation. *Development* **125**, 4293–4299 (1998).
- Li, Q., Van Antwerp, D., Mercurio, F., Lee, K. F. & Verma, I. M. Severe liver degeneration in mice lacking the I κ B kinase 2 gene. *Science* **284**, 321–325 (1999).
- Li, Z. W. *et al.* The IKK β subunit of I κ B kinase (IKK) is essential for nuclear factor κ B activation and prevention of apoptosis. *J. Exp. Med.* **189**, 1839–1845 (1999).
- Beg, A. A., Sha, W. C., Bronson, R. T., Ghosh, S. & Baltimore, D. Embryonic lethality and liver degeneration in mice lacking the RelA component of NF- κ B. *Nature* **376**, 167–170 (1995).
- Yeh, W. C. *et al.* Early lethality, functional NF- κ B activation, and increased sensitivity to TNF-induced cell death in TRAF2-deficient mice. *Immunity* **7**, 715–725 (1997).

- Beyaert, R., Vanhaesebroeck, B., Suffys, P., Van Roy, F. & Fiers, W. Lithium chloride potentiates tumor necrosis factor-mediated cytotoxicity *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* **86**, 9494–9498 (1989).
- Beg, A. A. & Baltimore, D. An essential role for NF- κ B in preventing TNF- α -induced cell death. *Science* **274**, 782–784 (1996).
- Van Antwerp, D. J., Martin, S. J., Kafri, T., Green, D. R. & Verma, I. M. Suppression of TNF- α -induced apoptosis by NF- κ B. *Science* **274**, 787–789 (1996).
- Wang, C. Y., Mayo, M. W. & Baldwin, A. S. TNF- and cancer therapy-induced apoptosis: potentiation by inhibition of NF- κ B. *Science* **274**, 784–787 (1996).
- Liu, Z. G., Hsu, H., Goeddel, D. V. & Karin, M. Dissection of TNF receptor 1 effector functions: JNK activation is not linked to apoptosis while NF- κ B activation prevents cell death. *Cell* **87**, 565–576 (1996).
- Baueuerle, P. A. & Baltimore, D. I κ B: a specific inhibitor of the NF- κ B transcription factor. *Science* **242**, 540–546 (1988).
- Oliver, F. J. *et al.* Resistance to endotoxic shock as a consequence of defective NF- κ B activation in poly (ADP-ribose) polymerase-1 deficient mice. *EMBO J.* **18**, 4446–4454 (1999).
- Beraud, C., Henzel, W. J. & Baeuerle, P. A. Involvement of regulatory and catalytic subunits of phosphoinositide 3-kinase in NF- κ B activation. *Proc. Natl Acad. Sci. USA* **96**, 429–434 (1999).
- Kane, L. P., Shapiro, V. S., Stokoe, E. & Weiss, A. Induction of NF- κ B by the Akt/PKB kinase. *Curr. Biol.* **9**, 601–604 (1999).
- Romashkova, J. A. & Makarov, S. S. NF- κ B is a target of AKT in anti-apoptotic PDGF signalling. *Nature* **401**, 86–90 (1999).
- Ozes, O. N. *et al.* NF- κ B activation by tumour necrosis factor requires the Akt serine-threonine kinase. *Nature* **401**, 82–85 (1999).
- Sizemore, N., Leung, S. & Stark, G. R. Activation of phosphatidylinositol 3-kinase in response to interleukin-1 leads to phosphorylation and activation of the NF- κ B p65/RelA subunit. *Mol. Cell. Biol.* **19**, 4798–4805 (1999).
- Cross, D. A., Alessi, D. R., Cohen, P., Andjelkovich, M. & Hemmings, B. A. Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **378**, 785–789 (1995).
- Wijmsman, J. H. *et al.* A new method to detect apoptosis in paraffin sections: *in situ* end-labeling of fragmented DNA. *J. Histochem. Cytochem.* **41**, 7–12 (1993).

Acknowledgements

We thank W.-C. Yeh for the TRAF2-deficient mouse EFs and C. Mirtsos, M. Bonnard, T. Nicklee, A. Ali, M. Parsons, T. Mak, D. Wakeham and A. Shahinian for technical help and advice. K.P.H. is supported by a Medical Research Council of Canada Studentship. J.R.W. is supported by grants from the Medical Research Council and Howard Hughes Medical Institute and is a Medical Research Council Senior Scientist.

Correspondence and requests for materials should be addressed to J.R.W. (e-mail: jwoodgett@uhnres.utoronto.ca).

Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth

Gefeng Zhu^{*†}, Paul T. Spellman[‡], Tom Voipe^{§||}, Patrick O. Brown[¶], David Botstein[‡], Trisha N. Davis^{*} & Bruce Futcher^{||}

^{*} Department of Biochemistry, University of Washington, Seattle, Washington 98195-7350, USA

[‡] Department of Genetics, Stanford University Medical Centre, Stanford, California 94306-5120, USA

[§] Graduate Program in Genetics, State University of New York, Stony Brook, New York 11794-5215, USA

^{||} Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

[¶] Department of Biochemistry, Stanford University Medical Centre, Stanford, California 94306-5428, USA

There are about 800 genes in *Saccharomyces cerevisiae* whose transcription is cell-cycle regulated^{1,2}. Some of these form clusters of co-regulated genes¹. The 'CLB2' cluster contains 33 genes whose transcription peaks early in mitosis, including *CLB1*, *CLB2*, *SWI5*, *ACE2*, *CDC5*, *CDC20* and other genes important for mitosis¹. Here we find that the genes in this cluster lose their cell cycle regulation in a mutant that lacks two forkhead transcription factors, *Fkh1* and *Fkh2*. *Fkh2* protein is associated with the promoters of *CLB2*,

[†] Present address: Mayo Clinic, Department of Immunology, Rochester, Minnesota 55906, USA.

SWI5 and other genes of the cluster. These results indicate that Fkh proteins are transcription factors for the CLB2 cluster. The *fkh1 fkh2* mutant also displays aberrant regulation of the 'SIC1' cluster¹, whose member genes are expressed in the M–G1 interval and are involved in mitotic exit. This aberrant regulation may be due to aberrant expression of the transcription factors Swi5 and Ace2, which are members of the CLB2 cluster and controllers of the SIC1 cluster. Thus, a cascade of transcription factors operates late in the cell cycle. Finally, the *fkh1 fkh2* mutant displays a constitutive pseudohyphal morphology, indicating that Fkh1 and Fkh2 may help control the switch to this mode of growth.

We determined the binding site for Fkh1 protein (Fig. 1). This site was similar to a motif found in front of the genes of the CLB2 cluster¹ (Fig. 1). This motif is the binding site for a transcription factor called 'SFF' (SWI five factor)^{3–5}, whose components have not been identified. Furthermore, transcription of *FKH1* and *FKH2* is regulated according to the cell cycle, with peak transcription during S phase¹, consistent with the idea that Fkh1 and Fkh2 might be involved in cell-cycle regulation.

To see whether Fkh1 and Fkh2 regulate genes of the CLB2 cluster, we constructed *fkh1* and *fkh2* single and double mutants. Neither single mutant had an obvious phenotype, but the double mutant had unusual cell morphology (see below). We examined expression of cell-cycle regulated genes in the *fkh1 fkh2* mutant. *CLN2*, whose expression is independent of SFF, displayed its normal late-G1 peak in these cells (Fig. 2). However, *SWI5*, an SFF-dependent gene⁴ and a member of the CLB2 cluster, failed to oscillate, but instead was constitutively expressed in moderate amounts (Figs 2 and 3). Interestingly, during α -factor arrest, *SWI5* was expressed in the *fkh1 fkh2* mutant but not in wild-type cells (not shown), indicating that Fkh1 and Fkh2 can repress as well as activate transcription.

We used microarrays for a more comprehensive analysis (Fig. 3). To examine cell cycle regulation, we compared synchronous $\Delta fkh1 \Delta fkh2$ cells to asynchronous $\Delta fkh1 \Delta fkh2$ cells. Although most genes were regulated normally in the *fkh1 fkh2* mutant after release from an α -factor block, the genes of the CLB2 cluster were an exception, and largely lost their cell cycle regulation. Of the 33 genes in the CLB2 cluster, 20 showed little or no oscillation in the *fkh1 fkh2* mutant (*ACE2*, *ALK1*, *BUD3*, *BUD4*, *CDC5*, *CLB1*, *CLB2*, *HST3*, *KIP2*, *IQG1*, *MOB1*, *MYO1*, *SWI5*, *YCL012w*, *YIL158w*, *YLR190w*, *YML033w*, *YML034w*, *YNL058c*, *YPL141c*), though they clearly oscillated in the wild type. Seven genes (*APC1*, *BUD8*, *NUM1*, *TEM1*, *YCL063w*, *YLR057w* and *YLR084c*) had little or no oscillation in the *fkh1 fkh2* mutant, but also had less than 2.5-fold oscillation in wild-type cells after release from an α -factor block, so their regulation by Fkh1 and Fkh2 is difficult to ascertain. The remaining six genes (*CDC20*, *CHS2*, *HOF1*, *YJL051w*, *YML119w* and *YPR156c*) retained a residual oscillation in the *fkh1 fkh2* mutant. Although the *fkh1 fkh2* mutations eliminated oscillation of the transcripts of the CLB2 cluster, moderate, constitutive expression remained for each gene (Fig. 3). Consistent with this, moderate, constitutive *CLB2*

expression is seen when the SFF sites are removed from the *CLB2* promoter^{5,6}. The constitutive expression of the genes in the CLB2 cluster explains why the *fkh1 fkh2* mutant is viable.

Misregulation of genes in the CLB2 cluster might result in secondary effects. In particular, the CLB2 cluster encodes the cell cycle transcription factors Swi5 and Ace2. These related factors^{7,8} are responsible for the M–G1 phase transcription of genes in the downstream 'SIC1' cluster^{1,8–10}. Indeed, genes in the SIC1 cluster were also misregulated in the *fkh1 fkh2* mutant (Fig. 3). Some genes had reduced expression (for example, *EGT2*, *CTS1*, *PCL9*); some genes had reduced oscillation (for example, *SIC1*, *YDL117w* and *PRY3*); and some genes had alterations in both time and amount of expression (for example, *YGL028c*). In the mutant, *SIC1* was expressed at the α -factor block, perhaps because *SWI5* is now expressed at the α -factor block. The diversity of responses may reflect different degrees of dependence on the amount of Swi5/Ace2.

Outside the CLB2 and SIC2 clusters, there were only a few genes whose regulation during the cell cycle was affected by the *fkh1 fkh2* mutation (for example, *BUD9*, *CLN3*, *YMR215w*, *KIN3* and *YOR315w*). Several genes had an overall expression that increased (for example, *YGP1*) or decreased (for example, *TAO3*, *YGL028c*, *YHR143w*, *SPS4*, *SUN4*) in asynchronous *fkh1 fkh2* cells compared with wild-type cells (that is, these quantitative changes did not necessarily involve altered cell-cycle periodicity). Some of the downregulated genes are involved in cell-wall metabolism or cell separation, and this may help explain the cell separation defect of *fkh1 fkh2* cells (see below). The genes showing the largest quantitative effects in asynchronous cells did not include any genes from the CLB2 cluster (see Supplementary Information and our web site: genome-www.stanford.edu/fkh), suggesting that these quantitative effects were indirect.

To distinguish direct and indirect effects, and show that Fkh proteins regulate the CLB2 cluster directly, we performed formaldehyde crosslinking immunoprecipitation¹¹. After immunoprecipitation of Fkh2 and associated chromatin, polymerase chain reaction (PCR) was used to test for various DNA fragments. Promoter fragments containing the SFF motifs from four genes of the CLB2 cluster (*SWI5*, *CLB2*, *YJL051w* and *HST3*) were assayed. All four were specifically present in the Fkh2 immunoprecipitate (Fig. 4a–c).

A	6	0	19	19	19	0	19	12	12	
T	0	15	0	0	0	1	0	5	7	
C	1	4	0	0	0	17	0	2	0	
G	12	0	0	0	0	1	0	0	0	
	R	Y	A	A	A	C	A	W	W	Consensus FKH1
	G	T	A	A	A	Y	A	A		Published SFF
	R	W	A	A	A	Y	A	W		SFF/Fkh site

Figure 1 The Fkh1 binding site. The Fkh1 site was determined using a GST–Fkh1 fusion and a modified oligonucleotide selection and amplification binding protocol (SAAB)²⁵. Nineteen oligonucleotides from the fifth cycle of SAAB were sequenced. The number of occurrences of each nucleotide at each position is shown. Published SFF is from ref. 1. SFF/FKH site is our best current estimate of the site. R, A or G; Y, T or C; W, A or T.

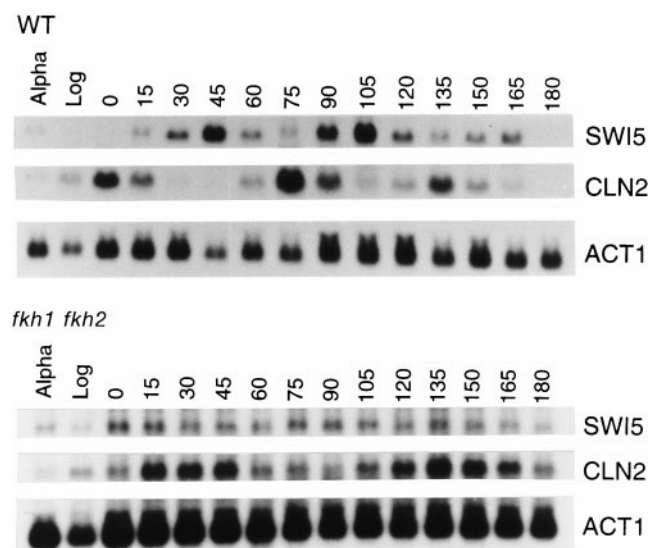


Figure 2 Regulation of *SWI5* during the cell cycle is lost in an *fkh1 fkh2* mutant. Wild-type (WT) or mutant cells were synchronized in G1 using α -factor, then released. *SWI5*, *CLN2*, and *ACT1* (loading control) messenger RNAs were assayed using northern blots. Time after release is shown (0–180 min). Alpha, α -factor arrested cells; Log, asynchronous cells.

In addition, we examined full-length promoters from three genes of the SIC1 cluster, *EGT2*, *SIC1* and *PCL9*. No fragment from the intergenic region upstream of these genes was specifically present in the immunoprecipitate (Fig. 4c–g, and data not shown for *PCL9*). Thus, Fkh2 directly regulates *SWI5*, *CLB2*, *YJL051w* and *HST3*, but only indirectly regulates *EGT2*, *SIC1* and *PCL9*.

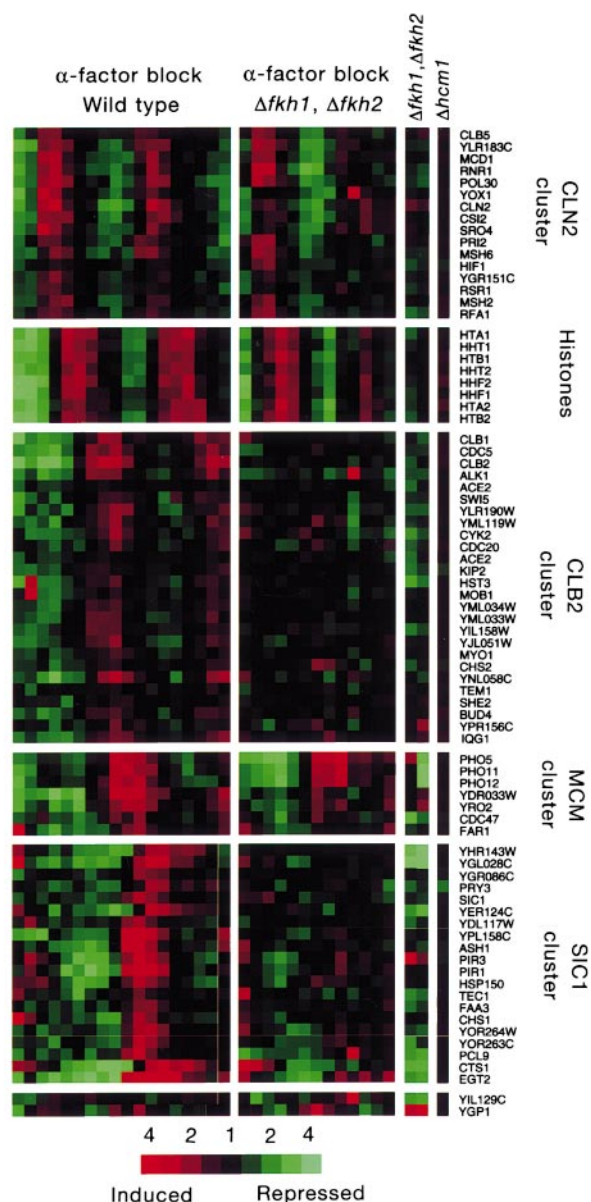


Figure 3 Microarray analysis of *fkh1 fkh2* mutants. Wild-type (left column) or isogenic *fkh1 fkh2* cells (GZ45-17a) (second column) were synchronized with α -factor, released, and sampled (left to right) through two cell cycles. Relative mRNA abundance was analysed by competitive microarray hybridization¹. Red, gene induction compared to asynchronous cells; green, repression; dynamic range is 16-fold. Representative genes from five clusters are shown. For α -factor experiments, synchronous wild-type cells were compared to asynchronous wild-type cells, and synchronous mutant cells were compared to asynchronous mutant cells, thus showing the effect of the mutations on oscillations in gene expression over the cell cycle. The ' $\Delta fkh1, \Delta fkh2$ ' column compares asynchronous *fkh1 fkh2* cells to asynchronous wild-type cells, and shows the effect of the mutations on overall expression. The ' $\Delta hcm1$ ' column compares asynchronous *hcm1* (a third forkhead-related gene) mutant cells to asynchronous wild-type cells. The bottom row shows the most upregulated gene in *fkh1 fkh2* mutants, *YGP1* (up 12-fold), and the most downregulated gene, *YIL129C* (down 30-fold). See Supplementary Information and our Web site (genome-www.stanford.edu/fkh) for complete data.

To summarize, almost all of the genes in the CLB2 cluster mostly or completely cease to oscillate in the *fkh1 fkh2* mutant, though they continue to be expressed. Many genes in the SIC1 cluster lose their normal regulation qualitatively and/or quantitatively, presumably because the transcription factors controlling them are encoded in the CLB2 cluster. Relatively few other genes are affected. We conclude that Fkh1 and Fkh2 are responsible for the regulation of the CLB2 cluster and probably encode components of SFF, because the Fkh1 binding site matches the SFF site found in the promoters of the CLB2 cluster, because SFF-regulated genes fail to oscillate in the *fkh1 fkh2* mutant, and because Fkh2 is actually present at four of these promoters. Our best estimate of the site consensus is RWAAAYAW. The slight residual periodic expression of some CLB2 cluster genes in the *fkh1 fkh2* mutant may be due to two related forkhead transcription factors, Hcm1 and Fhl1 (ref. 12).

The *fkh1 fkh2* mutants had striking morphological phenotypes:

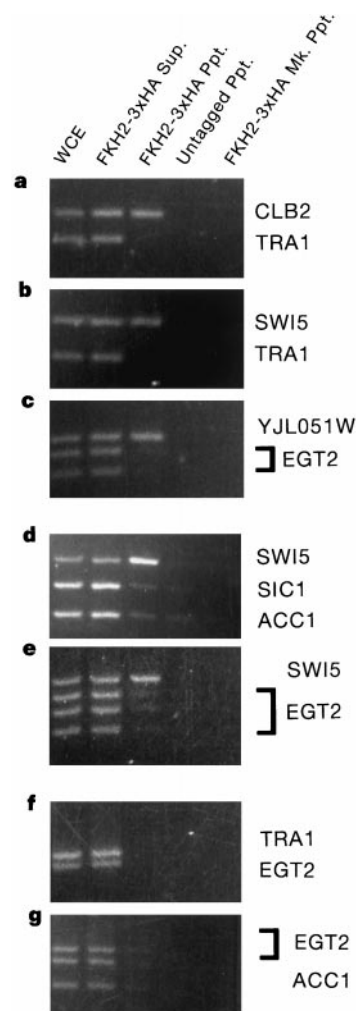


Figure 4 Fkh2 is at the promoters of *SWI5*, *CLB2* and *YJL051w*, but not at *EGT2* or *SIC1*. PCR fragments amplified from anti-Fkh2-3xHA immunoprecipitation or other chromatin fractions are shown after gel electrophoresis and ethidium bromide staining. WCE, whole cell extract; FKH2-3xHA Sup., supernatant from the immunoprecipitation of the haemagglutinin (HA)-tagged strain; FKH2-3xHA Ppt., material eluted from the immunoprecipitation of the HA-tagged strain; Untagged Ppt., material eluted from the immunoprecipitation of the untagged control strain; FKH2-3xHA Mk. Ppt., material eluted from the mock immunoprecipitation (no antibody) of the HA-tagged strain; CLB2, SWI5 and YJL051w, fragments from the promoters of CLB2 cluster genes; TRA1 and ACC1, negative control fragments; EGT2 (three different fragments) and SIC1, fragments encompassing the complete upstream regions of *EGT2* and *SIC1* genes of the SIC1 cluster.

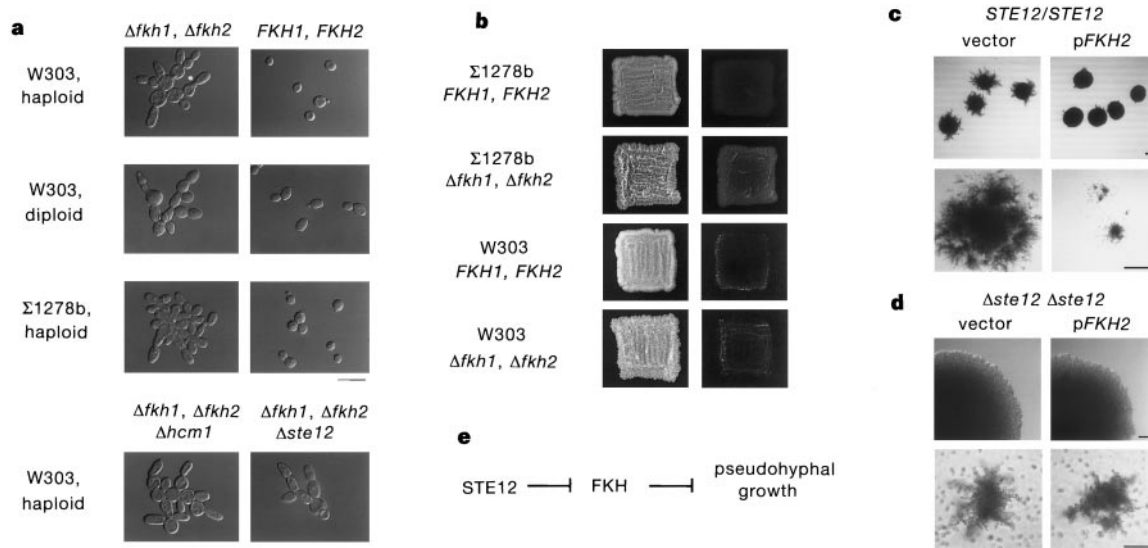


Figure 5 Phenotype of $\Delta fkh1 \Delta fkh2$ cells. **a**, Morphology. Yeast strains were grown in rich medium to mid-log phase, concentrated by centrifugation, sonicated and photographed. Scale bar: 10 μm . **b**, Invasiveness. Cells were patched onto rich medium and grown for two days ($\Sigma 1278b$ background) or three days (W303 background) at 30 °C. The first column shows the patches before washing, the second column shows the patches after washing with a stream of water. **c**, Extra copies of *FKH2* suppress formation

of filaments and invasion into the agar. A diploid strain from the $\Sigma 1278b$ background (L5366) was transformed with a multicopy control plasmid (pGF29) or a multicopy plasmid containing *FKH2* (pGF53), and grown on SLAD plates²⁶ at 30 °C. Top, colonies before washing; bottom, colonies after washing. Scale bars: 50 μm . **d**, As in **c** except the diploid from the $\Sigma 1278b$ background carried a homozygous deletion $\Delta ste12/\Delta ste12$.

mother and daughter cells remained attached; mother and daughter cells budded synchronously (by time-lapse photography, not shown); cells were elongated (Fig. 5a); and cells were somewhat invasive on agar plates (Fig. 5b). These phenotypes occurred in W303 haploids and diploids in nutrient-rich solid or liquid media. The morphology is characteristic of pseudohyphal growth, which usually occurs after nitrogen starvation on solid medium and allows yeast to forage more efficiently^{13–15}. However, pseudohyphal growth does not usually occur in strain W303. The phenotypes were not seen in *fkh1 hcm1* or *fkh2 hcm1* mutants, nor were they intensified in a *fkh1 fkh2 hcm1* triple mutant (Fig. 5a). It is consistent with our findings that *clb2* mutants themselves are weakly pseudohyphal¹⁶, as transcription of the CLB2 cluster is activated by Clb2/Cdc28 kinase activity¹⁷. A *Schizosaccharomyces pombe* mutant defective in a forkhead transcription factor, *sep1*, also has a cell separation defect¹⁸.

Unlike strain W303, strain $\Sigma 1278b$ undertakes robust pseudohyphal growth when starved for nitrogen on solid media¹³. The *fkh1 fkh2* mutations allowed pseudohyphal and invasive growth of $\Sigma 1278b$ even in rich media (Fig. 5a, b).

It is possible that the *fkh* mutants are only mimicking pseudohyphal growth (that is, they may be pseudo-pseudohyphal). However, overexpression of *FKH2* suppressed the normal ability of $\Sigma 1278b$ -related cells to become pseudohyphal upon nitrogen starvation (Fig. 5c), indicating that Fkh2 may be a part of the normal pathway for this adaptation.

Deletion of *STE12* dramatically reduces (but does not abolish) normal pseudohyphal growth¹⁴. In contrast, a *ste12 fkh1 fkh2* triple mutant is pseudohyphal, like the *fkh1 fkh2* double mutant (Fig. 5a). Thus, the Fkh proteins either act downstream of Ste12, or in a parallel pathway. However, overexpression of *FKH2* did not reduce the residual capacity of an *ste12* mutant for pseudohyphal growth (Fig. 5d), arguing that the Fkh proteins are not in a parallel pathway. In summary, Ste12 may control pseudohyphal growth in part by repressing Fkh expression or activity, which in turn may repress some aspects of pseudohyphal growth (Fig. 5e).

The pseudohyphal phenotypes can be explained in terms of genes

in the CLB2 and SIC1 clusters. The CLB2 cluster contains *BUD3*, 4 and 8, which affect budding pattern. The synchronous budding and elongated cells might be caused by a delay in mitosis, and many genes important for mitosis are in the CLB2 and SIC1 clusters. Finally, the lack of mother–daughter separation could be due to the decreased expression of the cell separation genes *EGT2* (ref. 9) and *CTS1* (ref. 19) of the SIC1 cluster, and perhaps also to the decreased expression of *YIL129c*, *YGL028c*, *YHR143w*, *SPS4*, *PRY3* and *SUN4*.

FKH1 and *FKH2* encode proteins homologous to the forkhead transcription factor of *Drosophila*²⁰. This family of transcription factors is highly conserved, with at least 20 homologues in humans^{21,22}. Here, we have shown that a pair of forkhead transcription factors is responsible for the M-phase transcription of a cluster of genes. As transcription of vertebrate B-type cyclins is also regulated during the cell cycle^{23,24}, it will be interesting to see whether any aspects of the regulatory system have been conserved. Our study of the *FKH* genes also helps understanding of the yeast cell cycle—the Fkh transcription factors are expressed in S phase to help to induce a set of genes in mitosis; this mitotic cluster includes the Swi5 and Ace2 transcription factors which then induce yet another cluster of genes during the M–G1 interval.

In developmental processes controlled by forkhead transcription factors in other species, it is not clear what the ultimate target genes may be; here, we have identified a number of direct and indirect targets. Similar microarray studies in other organisms may define the final target genes of other developmental pathways. □

Methods

Determination of the FKH1 binding site

The FKH1 site was determined using a modified selection and amplification binding protocol (SAAB)²⁵. See Supplementary Information and our Web site (genome-www.stanford.edu/fkh).

Analysis of gene expression

Strain W303a (*MATa ade2 leu2 his3 trp1 ura3 ssd1-d can1-100*) and isogenic GZY45-17a *MATa bar1 fkh1 fkh2* cells were synchronized using α -factor. Samples were taken every 15 min after release from α -factor. RNA was analysed by northern blotting using probes for

SWI5, *CLN2* and *ACT1*. Microarray analysis was as described¹. Microarray data for α -factor block-release of wild-type cells is from ref. 1. Strain GZY45-17a showed twofold increases in expression for many genes on chromosome 16, indicating that this strain could be disomic.

Crosslinking chromatin immunoprecipitations

Chromatin immunoprecipitations were carried out as described¹¹ with minor modifications. Exact methods and oligos are described in Supplementary Information and at our Web site (genome-www.stanford.edu/fkh).

Received 7 February; accepted 20 April 2000.

1. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
2. Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
3. Althoefer, H., Schleiffer, A., Wassmann, K., Nordheim, A. & Ammerer, G. Mcm1 is required to coordinate G2-specific transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 5917–5928 (1995).
4. Lydall, D., Ammerer, G. & Nasmyth, K. A new role for MCM1 in yeast: cell cycle regulation of *SWI5* transcription. *Genes Dev.* **5**, 2405–2419 (1991).
5. Maher, M., Cong, F., Kindelberger, D., Nasmyth, K. & Dalton, S. Cell cycle-regulated transcription of the *CLB2* gene is dependent on Mcm1 and a ternary complex factor. *Mol. Cell. Biol.* **15**, 3129–3137 (1995).
6. Seufert, W., Futcher, B. & Jentsch, S. Role of a ubiquitin-conjugating enzyme in degradation of S- and M-phase cyclins. *Nature* **373**, 78–81 (1995).
7. Moll, T., Tebb, G., Surana, U., Robitsch, H. & Nasmyth, K. The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor *SWI5*. *Cell* **66**, 743–758 (1991).
8. Dohrmann, P. R. *et al.* Parallel pathways of gene regulation: homologous regulators *SWI5* and *ACE2* differentially control transcription of HO and chitinase. *Genes Dev.* **6**, 93–104 (1992).
9. Kovacech, B., Nasmyth, K. & Schuster, T. EGT2 gene transcription is induced predominantly by *Swi5* in early G1. *Mol. Cell. Biol.* **16**, 3264–3274 (1996).
10. Knapp, D., Bhoite, L., Stillman, D. J. & Nasmyth, K. The transcription factor *Swi5* regulates expression of the cyclin kinase inhibitor *p40SIC1*. *Mol. Cell. Biol.* **16**, 5701–5707 (1996).
11. Hecht, A. & Grunstein, M. Mapping DNA interaction sites of chromosomal proteins using immunoprecipitation and polymerase chain reaction. *Methods Enzymol.* **304**, 399–314 (1999).
12. Zhu, G. & Davis, T. N. The fork head transcription factor Hcm1p participates in the regulation of SPC110, which encodes the calmodulin-binding protein in the yeast spindle pole body. *Biochim. Biophys. Acta* **1448**, 236–244 (1998).
13. Gimeno, C. J., Ljungdahl, P. O., Styles, C. A. & Fink, G. R. Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell* **68**, 1077–1090 (1992).
14. Liu, H., Styles, C. A. & Fink, G. R. Elements of the yeast pheromone response pathway required for filamentous growth of diploids. *Science* **262**, 1741–1744 (1993).
15. Kron, S. J., Styles, C. A. & Fink, G. R. Symmetric cell division in pseudohyphae of the yeast *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **5**, 1003–1022 (1994).
16. Edgington, N. P., Blacketer, M. J., Bierwagen, T. A. & Myers, A. M. Control of *Saccharomyces cerevisiae* filamentous growth by cyclin-dependent kinase Cdc28. *Mol. Cell. Biol.* **19**, 1369–1380 (1999).
17. Amon, A., Tyers, M., Futcher, B. & Nasmyth, K. Mechanisms that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins. *Cell* **74**, 993–1007 (1993).
18. Grallert, A., Grallert, B., Ribar, B. & Sipiczki, M. Coordination of initiation of nuclear division and initiation of cell division in *Schizosaccharomyces pombe*: genetic interactions of mutations. *J. Bacteriol.* **180**, 892–900 (1998).
19. King, L. & Butler, G. Ace2p, a regulator of CTS1 (chitinase) expression, affects pseudohyphal production in *Saccharomyces cerevisiae*. *Curr. Genet.* **34**, 183–191 (1998).
20. Weigel, D., Jurgens, G., Kuttner, F., Seifert, E. & Jackle, H. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. *Cell* **57**, 645–658 (1989).
21. Kaufmann, E. & Knochel, W. Five years on the wings of fork head. *Mech. Dev.* **57**, 3–20 (1996).
22. Lai, E., Clark, K. L., Burley, S. K. & Darnell, J. E. Hepatocyte nuclear factor 3/fork head or “winged helix” proteins: a family of transcription factors of diverse biologic function. *Proc. Natl Acad. Sci. USA* **90**, 10421–10423 (1993).
23. Cogswell, J. P., Godlevski, M. M., Bonham, M., Bisi, J. & Babiss, L. Upstream stimulatory factor regulates expression of the cell cycle-dependent cyclin B1 gene promoter. *Mol. Cell. Biol.* **15**, 2782–2790 (1995).
24. Hwang, A., Maity, A., McKenna, W. G. & Muschel, R. J. Cell cycle-dependent regulation of the cyclin B1 promoter. *J. Biol. Chem.* **270**, 28419–28424 (1995).
25. Overdier, D. G., Porcella, A. & Costa, R. H. The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid residues adjacent to the recognition helix. *Mol. Cell. Biol.* **14**, 2755–2766 (1994).
26. Gimeno, C. J. & Fink, G. R. Induction of pseudohyphal growth by overexpression of PHD1, a *Saccharomyces cerevisiae* gene related to transcriptional regulators of fungal development. *Mol. Cell. Biol.* **14**, 2100–2112 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank G. Fink for strains and H. Wijnen for reading the manuscript. This work was supported by grants from the NIH to T.D., D.B., P.B. and B.F.

Correspondence and requests for materials should be addressed to B.F. (e-mail: futcher@cshl.org).

Forkhead-like transcription factors recruit Ndd1 to the chromatin of G2/M-specific promoters

Manfred Koranda, Alexander Schleiffer*, Lukas Endler & Gustav Ammerer

Department of Biochemistry and Molecular Cell Biology, Ludwig Boltzmann Forschungsstelle, University of Vienna, Dr. Bohrgasse 9, A-1030 Vienna, Austria
* Present address: Institute of Molecular Pathology, Dr. Bohrgasse 7, A-1030 Vienna, Austria

Many cell-cycle-specific events are supported by stage-specific gene expression. In budding yeast, at least three different nuclear factors seem to cooperate in the periodic activation of G2/M-specific genes^{1–3}. Here we show, by using chromatin immunoprecipitation polymerase chain reaction assays, that a positive regulator, Ndd1, becomes associated with G2/M promoter regions in manner that depends on the stage in cell cycle. Its recruitment depends on a permanent protein–DNA complex consisting of the MADS box protein, Mcm1, and a recently identified partner Fkh2, a forkhead/winged helix related transcription factor^{4,5}. The lethality of Ndd1 depletion is suppressed by *fkh2* null mutations, which indicates that Fkh2 may also have a negative regulatory role in the transcription of G2/M-induced RNAs. We conclude that Ndd1–Fkh2 interactions may be the transcriptionally important process targeted by Cdk activity.

From the initiation of S phase to the completion of anaphase, yeast cells need to maintain a high level of B-type cyclin (Clb)-mediated Cdk1 activity. If Clb kinase activity decreases too much during this interval, then re-replication may occur before chromosome separation and cytokinesis⁶. One mechanism that ensures continuous production of the main mitotic cyclin, Clb2, is based on a positive feedback loop between the Clb2/Cdk1 kinase and the transcriptional activation system of *CLB2* (ref. 7). In G1 and early S phase, the messenger RNA level of this gene is low because of the lack of promoter activity^{2,3}. Re-accumulation of the mRNA during later stages of the cell cycle depends largely on Clb-dependent kinase activity. The machinery for *CLB2* activation

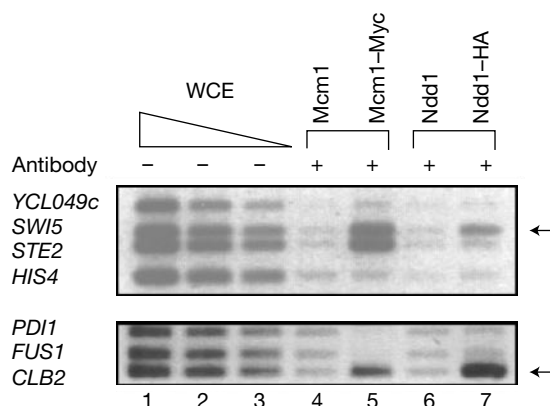


Figure 1 Ndd1 binds specifically to Mcm1-dependent G2/M-specific promoters. Chromatin immunoprecipitation PCR (ChIP) assays with *STE2*, *SWI5* and *CLB2* using tagged versions of Mcm1 and Ndd1. Lanes 1–3, dilution series of the whole-cell extract control (WCE); lanes 4 and 5, immunoprecipitation with anti-Myc antibody; lanes 6 and 7, immunoprecipitation with HA-specific antibody. Arrows emphasize the signals created by the *SWI5* and *CLB2* primer pairs. Gene names specify other promoter regions amplified by the control primer pairs. Top and bottom panels show results of independent experiments.

also coordinates the expression of a variety of factors needed during mitosis, exit from anaphase, cytokinesis and re-establishment of the next budding cycle⁸. Among these G2/M-specific genes is *SWI5*, whose product also functions as an important transcriptional regulator for cell-cycle-specific events taking place in late anaphase and during G1 (refs 9, 10). Indeed, the *SWI5* gene has served as the archetypical G2/M-specific transcriptional entity in yeast. Its cell-cycle regulatory pattern closely mimics the *CLB2* pattern, and its promoter has been extensively used to identify and characterize G2/M-specific regulatory constituents^{1,9}. The DNA sequences that impart this type of regulation recruit a protein complex comprising the MADS box protein Mcm1 and a partner called Sff (SWI five factor). Although only defined biochemically through its DNA-binding properties, the cooperation of Sff with Mcm1 seems to be necessary for G2/M-specific induction⁹.

NDD1 was originally identified as a candidate gene for Sff, but further characterization ruled this out³. We isolated *NDD1* as a high copy number suppressor that is able to overcome the G2/M-specific transcriptional defect of a special *mcm1* temperature-sensitive allele (A.S., unpublished data). We confirmed the requirement of Ndd1 for G2/M-specific transcription by constructing a galactose-dependent expression system for *NDD1*. A complete deletion of this gene was incompatible with either germination or cell survival; however, a plasmid containing a *GAL1* promoter/*NDD1* fusion allowed growth under de-repressed, galactose-induced conditions (see below). Depletion of Ndd1 under glucose-repressed conditions resulted in highly mishapen cells with 2C nuclei that exhibit low G2/M-specific gene expression (data not shown). All these observations are in agreement with previous results³.

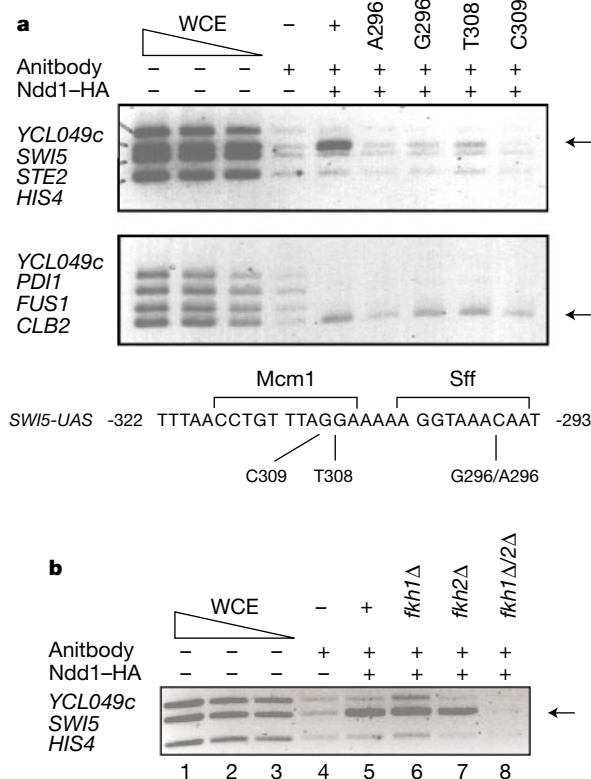


Figure 2 Ndd1 binding requires the ternary complex factor Sff. **a**, ChIP assays with a mixture of PCR primers including a *SWI5* primer pair (top panel) and *CLB2* primer pair (bottom panel). Extracts were generated from cells containing the wild-type *SWI5* promoter (lanes 4 and 5) or a point mutation at the specified position (lanes 6–9). The exact sequence of the *SWI5*-UAS mutant across the Mcm1–Sff binding sites and the nucleotide changes associated with the mutations are shown below. **b**, ChIP assays carried out on cells containing a deletion of *FKH1* (lane 6), *FKH2* (lane 7), or both (lane 8).

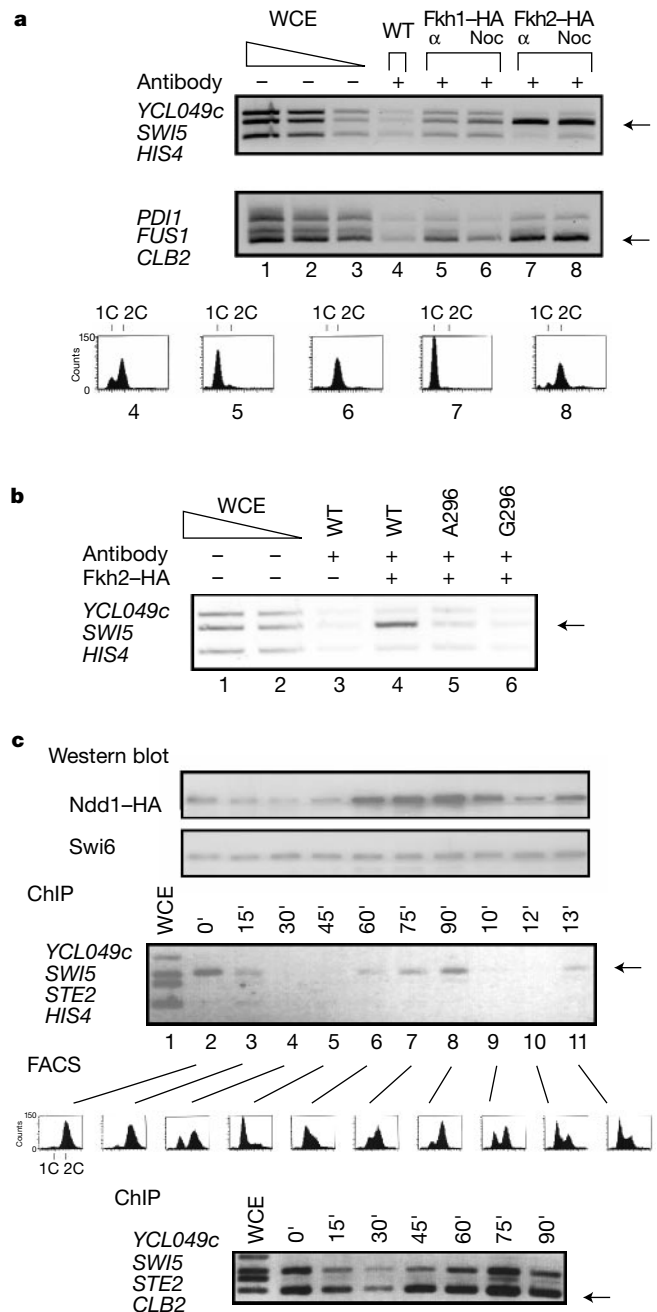


Figure 3 Fkh1 and Fkh2 binding to G2/M-specific promoters is independent of the stage in cell cycle, whereas Ndd1 is recruited to chromatin in a cell-cycle-specific manner. **a**, Strains containing either an HA-tagged version of the *FKH1* gene (lanes 5 and 6) or the *FKH2* gene (lanes 7 and 8) were used in a ChIP assay. Lane 4 is an untagged protein control. Cells were either treated with alpha mating pheromone (α , lanes 5 and 6) or nocodazole (noc, lanes 7 and 8). DNA profiles of the cultures derived from fluorescence-activated cell sorting (FACS) are given below with numbers corresponding to the lane numbers of the ChIP assay. **b**, The Fkh2-HA-dependent *SWI5* amplification signal depends on the integrity of the Sff-binding site. WT denotes precipitations from wild-type *SWI5* promoter cells, A296 and C296 (lanes 5 and 6) from cells containing the respective point mutation in the promoter. **c**, Extracts were prepared from *GAL1-CDC20 cdc20Δ* [pAS40] cells in G2 arrest and after carbon source release at the indicated times (in minutes). Samples were processed for western blot, ChIP assay and FACS as indicated.

Ndd1 has no extensive similarity to any proteins in public databases; however, it does show a weak relation to Rox1, an HMG-like nuclear repressor protein in yeast^{11,12}, although the exact biochemical function of this protein is also not well understood. Even though Ndd1 is found in the nucleus, attempts to show *in vitro* DNA binding at G2/M promoters have failed. Extracts from Ndd1-depleted cells show no defect in the formation of the ternary Mcm1–Sff complex on *SWI5* promoter fragments (ref. 3; and A.S., unpublished data). To test whether Ndd1 is part of a transcription regulatory complex that operates directly at G2/M promoters, we used chromatin immunoprecipitation polymerase chain reaction (ChIP) assays^{13,14}. A single copy plasmid expressing a haemagglutinin A (HA)-tagged version of *NDD1* was transformed into yeast, and we compared these transformants with cells expressing a Myc-tagged version of Mcm1 in ChIP assays (Fig. 1). As expected, Mcm1 is found at promoters that depend on this MADS box protein, such as the promoters of the pheromone receptor gene *STE2* (ref. 15) and two G2/M regulated genes *SWI5* and *CLB2* (ref. 1; Fig. 1, lane 5), whereas Ndd1–HA is specifically recruited to G2/M promoters. In contrast to results with Mcm1–Myc, we observed no Ndd1–HA binding to *STE2*, but a distinct and strong signal at the *SWI5* and *CLB2* promoters (Fig. 1, lane 7). Ndd1 is therefore a regulatory protein that functions directly at G2/M-specific promoters.

As Ndd1 has a strong preference for promoters containing an Mcm1–Sff binding motif, we determined whether the presence of

this complex is required for Ndd1 chromatin recruitment. Different point mutations introduced into the *SWI5* promoter⁹ highly affected the formation of an Mcm1–Sff complex *in vitro* and transcriptional performance *in vivo* (Fig. 2a). We tested four strains each containing a *SWI5*–UAS point mutation for Ndd1–HA binding. Although a weak signal can still be detected at one of the promoters containing a defective Mcm1-binding site (T308), signals are absent in both Sff-site mutants. Two forkhead-related proteins in yeast, Fkh1 and Fkh2, have recently been proposed as candidate genes encoding Sff^{4,5}. We therefore tested whether an *fkh1 fkh2* double mutant also exhibited loss of Ndd1–HA binding. The relevant PCR signal was absent in such strains, whereas it was retained in single mutants (Fig. 2b), indicating that both factors may be able to mediate Ndd1 recruitment. This result also supports the idea that either Fkh1 or Fkh2 is the binding partner of Mcm1 in G2/M transcription.

In vivo footprinting experiments suggested that Mcm1–Sff binding sites might be continuously occupied during the cell cycle¹, but whether the footprints were caused by the same or different cell-cycle-specific partners of Mcm1 was unresolved. We therefore tested if and when Fkh1 and Fkh2 can be found at G2/M-specific promoters. Strains containing functional, tagged versions of *FKH1* and *FKH2* were constructed and used for ChIP assays. To obtain cells in G1, we treated half of the culture with mating pheromone until most cells exhibited the relevant cytological

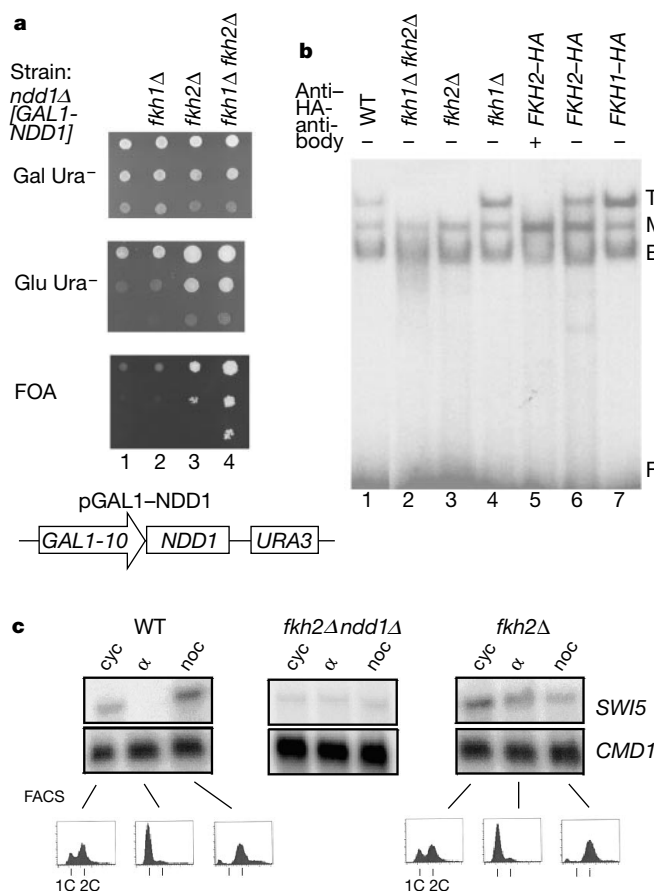


Figure 4 Fkh2 is the main constituent of the Sff-mediated cellular requirement for Ndd1. **a**, About 10^3 cells from logarithmically growing cultures were dropped on minus uracil plates containing galactose (top) or glucose (middle). Bottom panel shows the result from a plate containing fluoro orotic acid (FOA), a medium that kills cells expressing *URA3*. All spotted cells are derivatives of strain US262 (ref. 3) containing a chromosomal deletion of *NDD1* and the *URA3* plasmid pGAL1–NDD1. Deletions of *FKH1* and *FKH2* are as indicated (lanes 2–4). **b**, Gel-retardation assays with *SWI5* as probe and extracts from strains

containing either deleted or tagged *FKH1* and *FKH2* genes. The different Mcm1-containing complexes, T, M and B, correspond to those published in ref. 9. Formation of the T complex correlates with *SWI5* transcriptional activation, and its reconstitution has defined Sff activity. **c**, Northern analysis of wild-type, *fkh2* and *fkh2 ndd1* strains with *SWI5*-specific and *CMD1*-specific probes. Samples are from cycling cells (cyc), pheromone-arrested cells (α) and nocodazole-arrested cells (noc).

features. To obtain cells in G2, we treated the other half of the culture with nocodazole. The results indicated that both Fkh1–HA and Fkh2–HA can associate with *CLB2* and *SWI5* (Fig. 3a), and that their occupation rate does not vary significantly between cell-cycle stages. Overall, the signal generated by Fkh2–HA was much stronger than the signal generated by Fkh1–HA, especially with regard to *SWI5*. The assay did not allow a precise estimate about the binding stoichiometry, but suggested that Fkh2 has a more prominent role than Fkh1 in cell-cycle-specific transcription. The Fkh2–HA binding was also dependent on the integrity of the Sff consensus site, as strains containing the single nucleotide changes at position –296 lost this signal (Fig. 3b).

Cell-cycle-dependent Ndd1 binding was evaluated using cells arrested with high Clb2 kinase activity before sister chromatid separation by Cdc20 depletion using a carbon-source-regulated *CDC20* expression system¹⁴. Cells were synchronously released by Cdc20 induction. Strong binding could be observed in G2 but the signal was greatly diminished as cells entered G1. A strong Ndd1–HA chromatin binding signal reappeared during S phase (Fig. 3c). Western analysis indicated that fluctuating amounts of Ndd1 protein should be partially responsible for this effect (which in turn might be partially supported by weak cell-cycle-specific mRNA fluctuations), and our data thus corroborate a previous description of this phenomenon³. However, the Ndd1 signals found by western and ChIP analysis did not completely correlate in their intensity (Fig. 3c, compare lanes 2, 5, 6 and 8), indicating that mechanisms other than Ndd1 protein expression and stability may contribute to the formation of the chromatin complexes. Nevertheless, the timing of the G2/M transcriptional pattern seems to follow closely the peaks of chromatin association of Ndd1.

In the absence of Fkh1 and Fkh2, G2/M-specific transcripts mostly lose their cyclical expression pattern to attain a constant basal level of expression that is higher than the wild-type expression level observed in G1 cells^{4,5}. The increased basal level of G2/M-specific transcripts in *fkh1 fkh2* cells might explain why they remain viable in contrast to *ndd1* cells. Accordingly, the Mcm1–Sff complex might function as a platform for factors mediating positive and negative inputs, with Ndd1 antagonizing an Fkh-dependent repressive system. If such a model of Ndd1 action is correct, *fkh1 fkh2* mutants should suppress the lethality caused by Ndd1 depletion. This prediction was confirmed as *ndd1 fkh1 fkh2* mutants are viable (Fig. 4a). In control experiments with single *fkh* mutants, we discovered that *ndd1 fkh2* but not *ndd1 fkh1* strains are viable, supporting our idea that Fkh2 might play the predominant regulatory role at G2/M-specific promoters. Using gel-retardation assays with the *SWI5*-UAS mutant, we confirmed that Fkh2 alone is responsible for *in vitro* Sff activity (Fig. 4b). The biochemically defined Mcm1–Sff complex is absent in *fkh2* cells, and can only be supershifted when tagged Fkh2–HA is present in the extract. In summary, Fkh2 is not only the main forkhead protein associated with G2/M promoters but it may also be the main mediator of negative transcriptional regulation. To test this, we measured *SWI5* transcript levels in different mutant strains under conditions of cell cycle arrest in G1 or G2. *SWI5* mRNA levels were significantly lower in pheromone-treated wild-type cells (G1) than in cycling cells or nocodazole-arrested cells (G2) (Fig. 4c). This difference disappeared in *fkh2* or *fkh2 ndd1* cells. In agreement with a periodic repression/activation model the relative *SWI5* mRNA level was higher in cells with a functional *NDD1*.

Forkhead transcription factors have been implicated in many developmental processes¹⁶. They are necessary for proliferative responses and cell differentiation, and have been identified as targets of signal transduction systems. Our data establish their connection to a well-defined cell-cycle-specific program in yeast. In comparison with some other systems involving forkhead transcription factors, the function of the yeast factors Fkh1 and Fkh2 does not seem to be determined by nucleo-cytoplasmic shuttling¹⁷. Fkh1 and Fkh2 are

constitutively bound to promoters and seem to function by providing a permanent platform for further regulatory inputs such as Ndd1 recruitment. How does the mitotic Clb/Cdk kinase impact on this system? In principle, two mechanisms are converging on Ndd1. First, Ndd1 might be stabilized in a cell-cycle-specific manner either directly as a substrate of the kinase or indirectly through regulation of the ubiquitination/degradation machinery. Second, one could imagine that phosphorylation events control the interaction between Ndd1 and the Mcm1–Sff complex. Conclusive evidence for regulated phosphorylation of Ndd1 or Fkh2 has so far eluded us; notably, however, both Fkh1 and Fkh2 contain a so-called ‘FHA’ (forkhead-associated) domain that is thought to act as docking site for phospho-serine or phospho-threonine motifs¹⁸. □

Methods

Plasmids, yeast strains and growth conditions

For detection of Ndd1 with HA monoclonal antibodies, cells were transformed with plasmid pAS40. This plasmid is based on YCplac33 containing a genomic fragment spanning the *NDD1* open reading frame (ORF) and containing 750 base pairs (bp) of 5′-flanking sequence. An *SpeI* site was introduced by PCR before the stop codon of *NDD1*. An *XbaI* cassette with six HA-epitope tags was inserted into this site. This plasmid is able to complement an *ndd1-Δ* allele. The genomic ORF of *MCM1* was tagged at its carboxy terminus with a sequence encoding 18 Myc-epitope tags (YAS43) using a PCR-based strategy and the *Kluyveromyces fragilis* TRP1 gene as marker^{19,20}. *FKH1* and *FKH2* were modified using similar strategies with a 6×HA HisMX plasmid. The same PCR-based procedure was used to delete the region from 9 bp upstream to 449 bp downstream of the ORF of *FKH1* (MK88), 95 bp upstream to 411 bp downstream of the ORF of *FKH2* (MK89), and both ORFs (strain MK90). All primer sequences are available on request. All yeast strains were derivatives of the W303 background. The epitope taggings do not lead to any noticeable phenotypic effects. Strains K1927 (A296), K1928 (G296), K1929 (T308) and K1930 (C309) containing single nucleotide substitutions in the *SWI5* promoter⁹ were transformed or used in genetic crosses. Strain US262 containing the *ndd1::LEU2 pGAL-NDD1* (YCP50) (ref. 3) and strain K7428 *MATa cdc20::LEU2 GAL-CDC20* (ref. 14) have been described. For cell-cycle arrest, cells were treated either with alpha-factor for 2 h (3 μg ml⁻¹ final concentration) or the microtubule depolymerization drug nocodazole (30 μg ml⁻¹ final concentration) in YPD for 2 h.

DNA binding assays and other techniques

Chromatin immunoprecipitation PCR assays were performed as described^{13,14} with minor modifications. Yeast cells were crosslinked with 1% formaldehyde for 20 min at room temperature. The cell extracts were sonicated for 30 s twice and 20 s once using a Virsonic 100 cell disrupter (VirTis). The gels were photographed using Gel Doc 2000 (Biorad). The exact position and sequence of primer pairs are available on request. Detailed conditions for extract preparation and *in vitro* DNA-binding assays have been described⁹. Protein extracts preparation, western blot analysis, handling of antibodies and flow-cytometric DNA quantitation were performed as described²¹. The antibodies used were affinity purified anti-Myc mouse 9E11, serum containing rat mono clonal 3F10 and polyclonal rabbit antiserum against Swi6. Northern analysis with radioactively labelled *CMD1* and *SWI5* fragments as probes followed a common protocol^{1,3}.

Received 28 April; accepted 5 June 2000.

1. Althoefer, H., Schleiffer, A., Wassmann, K., Nordheim, A. & Ammerer, G. Mcm1 is required to coordinate G2-specific transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 5917–5928 (1995).
2. Maher, M., Cong, F., Kindelberger, D., Nasmyth, K. & Dalton, S. Cell cycle-regulated transcription of the *CLB2* gene is dependent on Mcm1 and a ternary complex factor. *Mol. Cell. Biol.* **15**, 3129–3137 (1995).
3. Loy, C. J., Lydall, D. & Surana, U. Ndd1, a high-dosage suppressor of *cdc28-1N*, is essential for expression of a subset of late-S-phase-specific genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 3312–3327 (1999).
4. Zhu, G. *et al.* Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature* **406**, 90–94 (2000).
5. Hollenhorst, P. C., Bose, M. E., Mielke, M. R., Müller, U. & Fox, C. A. Forkhead genes in transcriptional silencing, cell morphology and cell cycle: overlapping and distinct functions for FKH1 and FKH2 in *Saccharomyces cerevisiae*. *Genetics* **154**, 1533–1548 (2000).
6. Nasmyth, K. At the heart of the budding yeast cell cycle. *Trends Genet.* **12**, 405–412 (1996).
7. Amon, A., Tyers, M., Futcher, B. & Nasmyth, K. Mechanisms that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins. *Cell* **74**, 993–1007 (1993).
8. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
9. Lydall, D., Ammerer, G. & Nasmyth, K. A new role of Mcm1 in yeast: cell cycle regulation of *SWI5* transcription. *Genes Dev.* **5**, 2405–2419 (1991).
10. Knapp, D., Bhoite, L., Stillman, D. J. & Nasmyth, K. The transcription factor Swi5 regulates expression of the cyclin kinase inhibitor p40SIC1. *Mol. Cell. Biol.* **16**, 5701–5707 (1996).
11. Deckert, J., Rodriguez Torres, A. M., Simon, J. T. & Zitomer, R. S. Mutational analysis of Rox1, a DNA-binding repressor of hypoxic genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 6109–6117 (1995).
12. Di Flumeri, C., Liston, P., Acheson, N. H. & Keng, T. The HMG domain of the ROX1 protein mediates repression of HEM13 through overlapping DNA binding and oligomerization functions. *Nucleic Acids Res.* **24**, 808–815 (1996).

13. Hecht, A., Strahl-Bolsinger, S. & Grunstein, M. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* **383**, 92–96 (1996).
14. Cosma, M. P., Tanaka, T. & Nasmyth, K. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell* **97**, 299–311 (1999).
15. Errede, B. & Ammerer, G. STE12, a protein involved in cell-type-specific transcription and signal transduction in yeast, is part of protein-DNA complexes. *Genes Dev.* **3**, 1349–1361 (1989).
16. Kaufmann, E. & Knöchel, W. Five years on the wings of fork head. *Mech. Dev.* **57**, 3–20 (1996).
17. Biggs, W. H. et al. Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1. *Proc. Natl Acad. Sci. USA* **96**, 7421–7426 (1999).
18. Durocher, D., Henkel, J., Fersht, A. & Jackson, S. P. The FHA domain is a modular phosphopeptide recognition motif. *Mol. Cell* **4**, 387–394 (1999).
19. Knop, M. et al. Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15**, 963–972 (1999).
20. Zachariae, W., Shin, T. H., Galova, M., Obermaier, B. & Nasmyth, K. Identification of subunits of the anaphase-promoting complex of *Saccharomyces cerevisiae*. *Science* **274**, 1201–1204 (1996).
21. Piatti, S., Bohm, T., Cocker, J. H., Diffley, J. F. & Nasmyth, K. Activation of S-phase-promoting CDKs in late G1 defines a 'point of no return' after which Cdc6 synthesis cannot promote DNA replication in yeast. *Genes Dev.* **10**, 1516–1531 (1996).

Acknowledgements

We are especially grateful to B. Futcher for providing us with unpublished information about Fkh1 and Fkh2. We thank U. Surana, K. Nasmyth for plasmids and many helpful discussions; M. Galova for sharing materials; C. Zimmer for strain constructions; T. Tanaka and E. Bräbäck for help in establishing ChIP assays; A. Gartner and P. Alepuz for comments on the manuscript. This work was supported by the Austrian National Bank and by the Austrian Fonds zur Förderung wissenschaftlicher Forschung.

Correspondence and requests for materials should be addressed to G.A. (e-mail: ga@abc.univie.ac.at).

Elevated UV-B radiation reduces genome stability in plants

Gerhard Ries*, Werner Heller†, Holger Puchta*‡, Heinrich Sandermann†, Harald K. Seidlitz† & Barbara Hohn*

* Friedrich Miescher-Institut, PO Box 2543, CH-4002 Basel, Switzerland

† GSF-Forschungszentrum für Umwelt und Gesundheit, D-85764 Neuherberg, Germany

Long-term depletion of the stratospheric ozone layer contributes to an increase in terrestrial solar ultraviolet-B radiation^{1–3}. This has deleterious effects on living organisms, such as DNA damage^{4,5}. When exposed to elevated ultraviolet-B radiation (UV-B; 280–315 nm), plants display a wide variety of physiological and morphological responses characterized as acclimation and adaptation⁶. Here we show, using special sun simulators, that elevated solar UV-B doses increase the frequency of somatic homologous DNA rearrangements in *Arabidopsis* and tobacco plants. Increases in recombination are accompanied by a strong induction of photolyase and Rad51 gene expression. These genes are putatively involved in major DNA repair pathways, photo-reactivation and recombination repair^{7,8}. In mutant *Arabidopsis* plants that are deficient in photoreactivating ultraviolet-induced cyclobutane pyrimidine dimers, recombination under elevated UV-B regimes greatly exceeds wild-type levels. Our results show that homologous recombination repair pathways might be involved in eliminating UV-B-induced DNA lesions in plants. Thus, increases in terrestrial solar UV-B radiation as forecasted for the early 21st century may affect genome stability in plants.

The stratospheric ozone layer is the key factor in reducing solar UV-B radiation reaching the Earth's surface³. The continuous depletion of the ozone layer will, therefore, be accompanied by an increase in terrestrial UV-B irradiance, especially at wavelengths

below 300 nm^{1–3,5}. Whereas the impact of UV-B radiation on plant growth, development, physiology and morphology has been studied intensively⁹, little is known about its influence on plant genome stability. Homologous recombination is particularly important to plants as their genomes contain large amounts of repeated DNA sequences and highly homologous gene families¹⁰. In addition, rearrangements between homologous DNA sequences in somatic cells are strongly stimulated by DNA-damaging agents¹⁰.

To obtain a realistic estimate of the long-term impact of increased solar UV-B irradiance on plant genomes, we exposed *Arabidopsis* and tobacco plants over several generations to different levels of elevated UV-B radiation. Genome stability was monitored by assaying homologous recombination in treated plants and their offspring. To simulate the natural photobiological environment we provided a natural spectral balance of photosynthetic active radiation (PAR; 400–700 nm) of 800 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and ultraviolet-A radiation (UV-A; 315–400 nm) of 19 W m⁻², conditions under which photobiological processes are activated to normal physiological levels¹¹. The spectrum was supplemented with four different UV-B regimes¹² which are given as biologically effective (BE) weighted daily UV-B dose normalized at 300 nm¹³ (details on irradiation spectra and an alternative weighting are available as Supplementary Information). The 'ambient' UV-B regimes, UV-B1 (2.3 kJ m⁻² d⁻¹_{BE}) and UV-B2 (6.6 kJ m⁻² d⁻¹_{BE}), lie in the range of UV-B currently encountered during the growing period in northern mid-latitudes (40° N–60° N) at altitudes up to about 1,000 m above sea level. Two 'high' regimes, UV-B3 (18.6 kJ m⁻² d⁻¹_{BE}) and UV-B4 (27.1 kJ m⁻² d⁻¹_{BE}) were used to test the plant's response to strongly elevated UV-B radiation.

The influence of increased UV-B radiation on somatic recombination frequency was analysed using one transgenic *Nicotiana tabacum* line and four independent *Arabidopsis thaliana* lines, carrying one or more copies of a recombination reporter transgene at a single but different locus each. All recombination substrates contain two overlapping parts of the β -glucuronidase gene which are orientated as tandem or inverted repeats and are interrupted by a hygromycin resistance gene. Homologous recombination between these overlapping sequences leads to a complete, functional β -glucuronidase gene. The length of overlap between the two homo-

Table 1 Somatic recombination frequency per genome ($\times 10^{-6}$) in transgenic *Arabidopsis* and tobacco plants

Daily UV-B dose (kJ m ⁻² d ⁻¹) _{BE}	Line				
	N9	A11	A211	A231	A651
2.3 (UV-B1)	11.3 $\pm$ 2.1	35.3 $\pm$ 4.9	0.6 $\pm$ 0.1	0.1 $\pm$ 0.0	1.1 $\pm$ 0.1
6.6 (UV-B2)	52.3 $\pm$ 8.2	58.6 $\pm$ 6.2	2.8 $\pm$ 0.5	0.2 $\pm$ 0.1	2.2 $\pm$ 0.3
18.6 (UV-B3)	n.d.	70.9 $\pm$ 9.2	5.1 $\pm$ 0.7	0.5 $\pm$ 0.1	4.4 $\pm$ 0.5
27.1 (UV-B4)	83.5 $\pm$ 17.1	86.8 $\pm$ 13.3	8.4 $\pm$ 1.0	0.8 $\pm$ 0.2	6.9 $\pm$ 0.6
Daily UV-B dose (kJ m ⁻² d ⁻¹) _{BE}	Line				
	A87	A94			
2.3	0.5 $\pm$ 0.1	0.9 $\pm$ 0.1			
7.7	1.2 $\pm$ 0.2	6.5 $\pm$ 1.1			
13.4	2.2 $\pm$ 0.4	25.1 $\pm$ 2.9			
	Generation (Line A651)				
	F ₀	F ₁	F ₂		
2.3	1.3 $\pm$ 0.2	1.3 $\pm$ 0.2	1.4 $\pm$ 0.3		
6.6	2.4 $\pm$ 0.3	5.7 $\pm$ 0.8	7.7 $\pm$ 1.1		
18.6	4.9 $\pm$ 0.7	12.3 $\pm$ 1.8	15.7 $\pm$ 2.2		

Each value represents the mean of 80–120 plants $\pm$ s.e. The structure of the recombination substrates for the different lines has been described^{14,15}. In brief, the tobacco line N9 carries four copies of the recombination substrate in indirect orientation, with 566-base pair (bp) homologous overlapping sequence; the *Arabidopsis* line A11 contains one copy of the recombination substrate in direct orientation, with 1,213-bp overlap; the *Arabidopsis* lines A211 (one copy) and A231 (two copies) carry the recombination substrates in direct orientation with 566-bp overlap (P. Swoboda, personal communication), and line A651 one copy in indirect orientation, with 566-bp overlap. n.d., not determined.

‡ Present address: IPK, Institut für Pflanzen-genetik und Kulturpflanzenforschung, D-06466 Gatersleben, Germany.

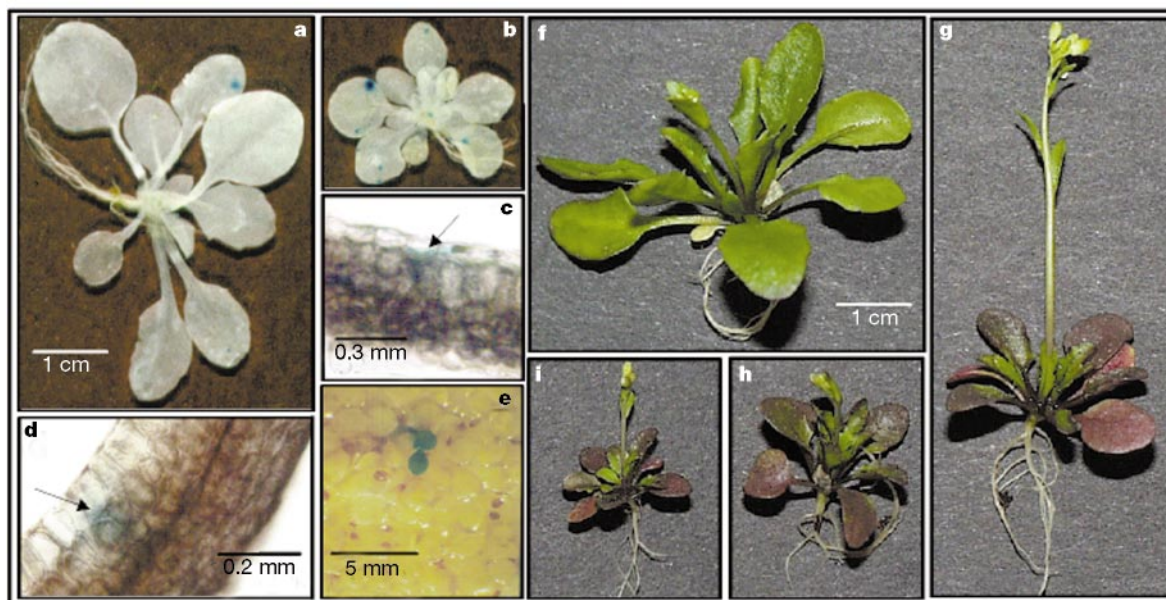


Figure 1 Transgenic *Arabidopsis thaliana* plants (line A11) grown under different daily UV-B doses. **a, b**, Visualization by histochemical staining of somatic recombination events in 23-day-old plants exposed to UV-B1 (**a**) and UV-B4 (**b**). The scale is the same for both. **c, d**, Sections through stained leaves from plants exposed to UV-B4. Arrows indicate recombination events. **e**, Germinal recombination event in an F_1 population of 7-day-old

seedlings grown under UV-B3 regime. The F_0 generation was grown under the identical UV-B3 regime. **f–i**, 23-day-old *Arabidopsis*, germinated and grown under $41 \text{ mol m}^{-2} \text{ d}^{-1}$ photosynthetic active radiation with different daily UV-B doses. **f**, UV-B1 = $2.3 \text{ kJ m}^{-2} \text{ d}^{-1}$; **g**, UV-B2 = $6.6 \text{ kJ m}^{-2} \text{ d}^{-1}$; **h**, UV-B3 = $18.6 \text{ kJ m}^{-2} \text{ d}^{-1}$; and **i**, UV-B4 = $27.1 \text{ kJ m}^{-2} \text{ d}^{-1}$. Scale is the same for **f–i**.

logous sections and their orientation to each other varies between recombination substrates^{14,15} (Table 1). After histochemical staining of plants, recombination events can be visualized and their number and location determined (Fig. 1a–e).

In all the lines tested, the somatic recombination frequency under the current ambient UV-B1 level ranged from $(1.0 \pm 0.2) \times 10^{-7}$ to $(3.5 \pm 0.5) \times 10^{-5}$ (mean $\pm$ s.e.) recombination events per genome (Table 1). The large variations between the lines may result from differences in the structure of the recombination substrates, transgene copy number and plant species. Differences in the genomic position of the randomly integrated recombination substrates may also be an important factor influencing both transcriptional activity and recombination frequency^{15,16}.

Exposure to different levels of elevated UV-B radiation induced morphological changes (Fig. 1f–i) and markedly affected the somatic recombination frequency in all the plant lines tested (Table 1). Under the elevated but still ecologically relevant UV-B2 regime the recombination frequency in tobacco and all *Arabidopsis* lines increased significantly ($P < 0.05$, t -test). The recombination frequencies in line N9 rose 4.6-fold, in line A11 1.7-fold, in line A211 4.6-fold and in lines A231 and A651 twofold. With the higher UV-B doses, UV-B3 and UV-B4, recombination was affected even more (up to 8.5-fold increases between UV-B1 and UV-B3 and up to 14-fold increases between UV-B1 and UV-B4). It seems that, in general, the induction of recombination under elevated UV-B radiation is independent of the plant species and line analysed and is probably a genome-wide phenomenon with varying specificity for different loci. The formation and position of UV-induced damage in DNA, as well as the accessibility and activity of DNA repair processes, are influenced largely by chromatin structure, at

least in yeast and mammalian cells^{17,18}. Different loci in the plant genome may also respond to UV-B stress with different increases in recombination activity.

Increased levels of recombination under elevated UV-B radiation may be an immediate consequence of increased damage and repair of affected DNA. To test whether repair of UV-B-induced DNA damage influences recombination, we crossed the transgenic *Arabidopsis* line A651 with the photolyase deficient line *uvr2-1* (ref. 19), which lacks photoreactivation repair of cyclobutane pyrimidine dimers, the major UV-B-induced DNA damage. In the resulting progeny line A94, homozygous for *uvr2-1*, the presence of UV-B radiation generally stimulated homologous recombination to a greater extent than in the control progeny line A87, homozygous for *UVR2-1* (Table 1). In UV-B1, the recombination frequency in line A94 was already about 1.8-fold higher than in line A87. Photolyase-deficient plants were much more sensitive to UV-B radiation than wild-type plants, therefore, we applied slightly reduced regimes. A daily dose of $7.7 \text{ kJ m}^{-2} \text{ d}^{-1}$ UV-B induced recombination in line A94 5.4 times more strongly than in line A87 and a dose of $13.4 \text{ kJ m}^{-2} \text{ d}^{-1}$ UV-B increased the recombination frequency up to 11.4 times more efficiently than in line A87. Higher UV-B doses were lethal for photolyase-deficient plants. The absence of cyclobutane pyrimidine dimer-specific photoreactivation repair increased the steady-state concentration of cyclobutane pyrimidine dimers in mutants exposed to elevated UV-B levels, compared with wild-type plants (Fig. 2a). In contrast, the concentration of (6-4) photoproducts was unaffected under elevated UV-B regimes in photolyase mutant plants, as in wild-type plants (data not shown). These data indicate that there may be a strong relationship between cyclobutane pyrimidine dimers in the plants' genomes and homologous recombination, and furthermore that recombination might be involved in cyclobutane pyrimidine dimer repair. Elimination of UV-B-induced DNA lesions, by DNA repair pathways such as nucleotide excision, may create recombinogenic intermediates. Alternatively, or in addition, UV-B-promoted signalling pathways may activate recombination processes by stimulating, for example, transcription of DNA repair genes²⁰ (see below).

Table 2 Germinal recombination frequency in *Arabidopsis* plants of line A11

Daily UV-B dose ($\text{kJ m}^{-2} \text{ d}^{-1}$) _{BE}	Germinal recombination (events per 250,000 seedlings)
2.3 (experiment 1)	1
2.3 (experiment 2)	2
6.6	4
18.6	10

Plants may compensate for deleterious effects of elevated UV-B radiation by mounting various acclimation responses, including accumulation of secondary metabolites and stimulation of DNA repair⁶. The accumulation of UV-B-screening compounds, mainly flavonols, is involved in protecting against UV-B radiation^{21,22}. We therefore determined the concentration of quercetin and kaempferol derivatives (Fig. 2b) as well as sinapate esters and anthocyanin compounds in UV-B-treated plant populations. In plants exposed to the UV-B2 level, the production of flavonols was significantly induced ($P < 0.01$, t -test) compared with UV-B1-treated plants. Under UV-B3 or UV-B4 conditions, only a weak additional flavonol accumulation could be observed (Fig. 2b). Unlike flavonols, sinapate esters were not induced by elevated UV-B radiation (data not shown). The content of the visible anthocyanins was increased under higher UV-B levels (Fig. 1f–i), although to much lower levels than those of flavonols (data not shown), indicating a minor role for the anthocyanins in absorbing UV-B. Under the anticipated elevated UV-B levels the protection against UV-B radiation afforded by UV-B-screening pigments may therefore be incomplete.

This may, at least in part, be counterbalanced by enhanced repair activity. We compared the repair activities in plants exposed to the different UV-B regimes by examining the steady-state levels of AtPhotolyase II (PHR1) and AtRad51 messenger RNA, the latter as a putative marker for recombination repair. Using polymerase chain reaction with reverse transcriptase (RT-PCR), we found that the steady-state mRNAs of both genes were increased under high UV-B doses (Fig. 2c). This suggests that plants grown under elevated UV-B regimes have an increased ability to repair DNA damage. This ability may reduce at least DNA damage resulting in cyclobutane pyrimidine dimers and (6-4) photoproducts to the level of that found in plants exposed to ambient UV-B conditions (Fig. 2a). In animals and yeast, Rad51 is crucial for recombination^{8,23} and in plant cells it has been reported to be strongly induced after gamma irradiation²⁴. If AtRad51 is involved in recombination

repair in plants, enhanced expression of AtRad51 under elevated UV-B levels could result in higher recombination repair activity, which in turn may contribute to increased frequencies of homologous recombination.

Germinal recombination events can permanently alter the genetic composition of subsequent generations and hence contribute to genomic changes in plant populations²⁵. Although reproductive organs are generally considered to be well protected from UV-B during both developmental and mature phases, long-term exposure to high UV-B levels may also affect the reproductive tissue (L2 cell layer) of plants and cause DNA damage. Mature pollen grains are potentially very susceptible to UV-B-induced DNA damage during the short period between anther dehiscence and pollen tube penetration into stigma tissues²⁶. Indeed, UV-B irradiation of maize pollen can activate cryptic transposable elements²⁷. Recombination processes may also be induced under such conditions, affecting the genome stability of future plant generations. We analysed the influence of UV-B radiation on the frequency of germinal recombination in *Arabidopsis* plants. Germinal recombinants are relatively rare in plants and can be the products of meiotic recombination or representatives of somatic homologous interactions¹⁶. To detect such rare germinal recombinants, we screened large numbers of seedlings of the subsequent generation of plants grown under elevated UV-B regimes. Compared to UV-B1 conditions, UV-B2 regimes increased the number of germinal recombinations in seven-day-old *Arabidopsis* seedlings about four times and UV-B3 regimes up to ten times (Fig. 1e, Table 2). The germination rate and the seed set did not differ between the analysed plant populations (data not shown). Although the numbers of germinal recombination events were small, this preliminary experiment may indicate that elevated UV-B levels can lead not only to an increase in genomic instability of somatic tissue but also to permanent changes in plant populations. Furthermore, the progeny of plants exposed to elevated UV-B exhibited a higher UV-B-induced somatic recombination rate than the parental population. Whereas ambient UV-B1 levels did not stimulate somatic recombination in subsequent plant generations, exposure of populations to UV-B2 and UV-B3 resulted in a marked increase in the recombination frequency in progeny plants (Table 1). When exposed to similar UV-B2 levels as the parental plants, the F_1 progeny showed a 2.4-fold and the F_2 progeny a 3.2-fold higher recombination frequency than F_0 plants. Under UV-B3 regimes the recombination frequencies in the F_1 and the F_2 generation increased by nearly the same factors of 2.5 and 3.2, respectively. Thus, elevated UV-B exposure over several generations may lead to progressive increases in somatic recombination rates as well as to higher numbers of permanently altered plants.

Under natural conditions the terrestrial UV-B irradiance varies strongly depending on the thickness of the UV-screening ozone layer, the solar elevation and meteorological conditions. Therefore, extrapolations from data obtained under experimental, static UV-B conditions to natural, dynamic UV-B field conditions are difficult. Our data, however, allow extrapolations on the entire ecologically relevant UV-B range because all recombination data obtained under the 'high' UV-B regimes (UV-B3, UV-B4) are in line with the results obtained under the currently ecological UV-B range (UV-B1, UV-B2). We believe, therefore, that the increase in solar UV-B radiation, due to a long-term depletion of the stratospheric ozone layer, may influence the genomic stability of plant populations. □

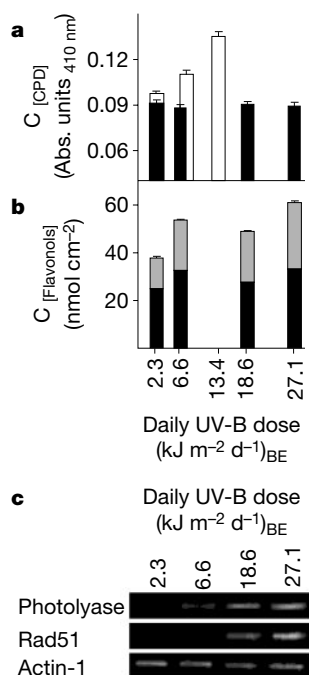


Figure 2 UV-B acclimation responses in *Arabidopsis thaliana* line A651 grown under different daily UV-B doses. **a**, Cyclobutane pyrimidine dimer (CPD) concentrations in wild-type (black bars) and photolyase-deficient plants (white bars). Vertical bars represent s.e. **b**, Flavonol concentrations, measured by quercetin (grey bars) and kaempferol (black bars) derivatives. Vertical bars represent s.d. **c**, Steady state mRNA levels for photolyase II (PHR1), AtRad51 protein and AtActin-1 determined by RT-PCR.

Methods

UV spectra

A combination of four lamp types (metal halide lamps, quartz halogen lamps, blue fluorescent tubes and UV-B fluorescent tubes) was used to obtain a natural balance of simulated global radiation throughout the ultraviolet to infrared spectrum. The short-wave cut-off was shaped by selected borosilicate, soda-lime and acrylic glass sheets¹². The mean PAR was 800 $\mu\text{mol m}^{-2} \text{s}^{-1}$, which corresponds to a daily dose of

~41 mol m⁻² d⁻¹ PAR. The calculations for the UV-B doses were performed with the general plant action spectrum normalized at 300 nm¹³. For comparison, we also provide the UV levels weighted with the action spectrum for dimer formation also normalised at 300 nm (dimer)²⁸ (for details see Supplementary Information).¹¹ The ratio (UV-B:PAR) is based on the non-weighted irradiance in the PAR and UV-B ranges. The spectroradiometric measurements were performed with a double monochromator system¹². Variations of the integrated values were less than 15%. We kept the UV-A irradiance constant at ~19 W m⁻² (non-weighted value) which corresponds to the moderate PAR value. Under the UV-B1 regime, the simulated UV-B:UV-A:PAR ratio of 1:50:480 (rounded figures) closely approximates an average present-day outdoor value around noon of 1:47:400 in northern mid-latitudes under an ozone column of ~320 Dobson units (typical ratio measured at Neuherberg (48° 41' N, 11° 41' E), South Germany, 495 m above sea level).

Growth conditions

Plants were germinated in soil and grown under 800 µmol m⁻² s⁻¹ PAR and different UV-B regimes. All climate parameters (14 h light at 20 °C, 10 h dark at 16 °C, 70% humidity) were identical for all ultraviolet treatments.

Somatic recombination frequency

23-day-old transgenic plants were stained histochemically and the recombination frequencies per genome were calculated¹⁴. In photolyase-deficient plants the recombination frequency was analysed 18 days after germination.

Germinal recombination frequency

Germinal recombination was determined in *Arabidopsis* seedlings of line A11. The F₀ generation was germinated and grown under UV-B1, UV-B2 or UV-B3 until seeds could be harvested. For each UV-B condition around 250,000 seeds were sown on wet filter paper placed on soil. After seven days exposure to similar UV-B regimes as progenitor plants, the seedlings were histochemically stained and the number of fully stained seedlings determined.

Crossings

Arabidopsis photolyase mutant (*uvr2-1*)¹⁹ was crossed with line A651 and independent lines were generated. F₃-generation lines A94 (homozygous for *uvr2-1*) and A87 (homozygous for *UVR2-1*) are both homozygous for line A651 recombination substrate.

Dimer concentration

CPDs and (6-4) photoproducts were quantified²⁹ in 23-day-old wild-type plants and 18-day-old photolyase-deficient plants. DNA was extracted from 100 mg plant tissue. CPD- and (6-4) photoproduct-specific monoclonal antibodies were purchased from Kamiya Biomedical Company. Samples were measured in quadruplicate and compared by analysis of variance within each experiment, using the computer program STATISTICA.

Flavonol concentration

Quercetin and kaempferol derivatives were analysed³⁰ in 23-day-old plants as used in Fig. 2a. Samples were analysed in duplicate or triplicate within each experiment.

Steady state mRNA levels

We used ~100 mg of 23-day-old plants, which were germinated and grown under the indicated conditions, to prepare total RNA using standard protocols. Reverse transcription was performed according to the protocol of the manufacturer (Pharmacia) for RT beads using 1 µg of total RNA. The complementary DNA was diluted 1:100, and 1 µl was used for PCR with specific primers. The primer sequences were for AtPhotolyase II (*PHR1*) (AF053365) 5'-GCCGTCGTTTCAATCTGTT-3' and 3'-AATCGATTGGTGGTGAAGC-5'; for AtRad51 (U43528) 5'-CGTTGAGGAAGAGCA-3' and 3'-GTGGCCAAAACATC AATCC-5'; and for AtActin-1 (M20016) 5'-AAAGGATGCTTATGTTGGCG-3' and 3'-AGCCACATACATAGCAGGGG-5'. Products with predicted sizes of 103 base pairs (bp) for AtPhotolyase, 132 bp for AtRad51 and 259 bp for AtActin-1 were obtained. The actin transcript was used as constitutive control. The cDNAs for all tested genes were amplified using hot start PCR under the following conditions: 94 °C for 5 min; followed by 40 cycles for AtRad51, 35 cycles for photolyase and 30 cycles for AtActin-1 of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 90 s; ending with 10 min at 72 °C. Equal loading of each amplified gene sequence was determined by the control AtActin-1 PCR product.

Received 21 January; accepted 10 May 2000.

- Kerr, J. & McElroy, C. Evidence for large upward trends of ultraviolet-B radiation linked to ozone depletion. *Science* **262**, 1032–1034 (1993).
- Madronich, S., McKenzie, R. L., Björn, L. O. & Caldwell, M. M. Changes in biologically active ultraviolet radiation reaching the earth's surface. *J. Photochem. Photobiol. B* **46**, 5–19 (1998).
- McKenzie, R., Connor, B. & Bodeker, G. Increased summertime UV radiation in New Zealand in response to ozone loss. *Science* **285**, 1709–1711 (1999).
- Rozema, J., van de Staaij, J., Björn, L. O. & Caldwell, M. M. UV-B as an environmental factor in plant life: stress and regulation. *Trends Ecol. Evol.* **12**, 22–28 (1997).
- Rousseaux, M. C. *et al.* Ozone depletion and UVB radiation: Impact on plant DNA damage in southern South America. *Proc. Natl Acad. Sci. USA* **96**, 15310–15315 (1999).
- Jansen, M. A. K., Gaba, V. & Greenberg, B. M. Higher plants and UV-B radiation: balancing damage, repair and acclimation. *Trends Plant Sci.* **3**, 131–135 (1998).
- Britt, A. B. Molecular genetics of DNA repair in higher plants. *Trends Plant Sci.* **4**, 20–24 (1999).
- Vispe, S., Cazaux, C., Lesca, C. & Defais, M. Overexpression of Rad51 protein stimulates homologous

- recombination and increases resistance of mammalian cells to ionizing radiation. *Nucleic Acids Res.* **26**, 2859–2864 (1998).
- Tevini, M. in *UV-B Radiation and Ozone Depletion: Effects on Humans, Animals and Plants, Microorganisms and Materials* (ed. Tevini, M.) 125–153 (CRC Press, Boca Raton: Lewis Publishers, USA, 1993).
- Puchta, H. & Hohn, B. From centiMorgans to base pairs: homologous recombination in plants. *Trends Plant Sci.* **1**, 340–348 (1996).
- Caldwell, M. M., Flint, S. & Searles, P. S. Spectral balance and UV-B sensitivity of soybean: a field experiment. *Plant Cell Environ.* **17**, 267–276 (1994).
- Döhrring, T., Köfferlein, M., Thiel, S. & Seidlitz, H. K. Spectral shaping of artificial UV-B irradiation for vegetation stress research. *J. Plant Physiol.* **148**, 115–119 (1996).
- Caldwell, M. M. in *Photophysiology* Vol. 6 (ed. Giese, A. C.) 131–177 (Academic, New York, 1971).
- Swoboda, P., Gal, S., Hohn, B. & Puchta, H. Intrachromosomal homologous recombination in whole plants. *EMBO J.* **13**, 484–489 (1994).
- Puchta, H., Swoboda, P., Gal, S., Blot, M. & Hohn, B. Somatic intrachromosomal homologous recombination events in populations of plant siblings. *Plant Mol. Biol.* **28**, 281–292 (1995).
- Tovar, J. & Lichtenstein, C. Somatic and meiotic chromosomal recombination between inverted duplications in transgenic tobacco plants. *Plant Cell* **4**, 319–332 (1992).
- Gale, J. M. & Smerdon, M. J. UV induced (6-4) photoproducts are distributed differently than cyclobutane dimers in nucleosomes. *Photochem. Photobiol.* **51**, 411–417 (1990).
- Suter, B., Livingstone-Zatchej, M. & Thomas, F. Chromatin structure modulates DNA repair by photolysis *in vivo*. *EMBO J.* **16**, 2150–2160 (1997).
- Landry, L. G. *et al.* An *Arabidopsis* photolyase mutant is hypersensitive to ultraviolet-B radiation. *Proc. Natl Acad. Sci. USA* **94**, 328–332 (1997).
- Mount, D. W. Reprogramming transcription. *Nature* **383**, 763–764 (1996).
- Landry, L. G. & Walbot, V. Flavonoids can protect maize DNA from the induction of ultraviolet radiation damage. *Plant Physiol.* **105**, 881–889 (1994).
- Landry, L. G., Chapple, C. C. & Last, R. L. *Arabidopsis* mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. *Plant Physiol.* **109**, 1159–1166 (1995).
- Shinohara, A., Ogawa, H. & Ogawa, T. Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **69**, 457–470 (1992).
- Doutriaux, M. P., Couteau, F., Bergounioux, C. & White, C. Isolation and characterisation of the RAD51 and DMC1 homologs from *Arabidopsis thaliana*. *Mol. Gen. Genet.* **257**, 283–291 (1998).
- Walbot, V. On the life strategies of plants and animals. *Trends Genet.* **1**, 165–169 (1985).
- Jackson, J. F. DNA repair in pollen. A review. *Mutat. Res.* **181**, 17–29 (1987).
- Walbot, V. UV-B damage amplified by transposons in maize. *Nature* **397**, 398–399 (1999).
- Quaite, F. E., Sutherland, B. M. & Sutherland, J. C. Action spectrum for DNA damage in alfalfa lowers predicted impact of ozone depletion. *Nature* **358**, 576–578 (1992).
- Stapleton, A. E., Mori, T. & Walbot, V. A simple and sensitive antibody-based method to measure UV-induced DNA damage in *Zea mays*. *Plant Mol. Biol. Reporter* **11**, 230–236 (1993).
- Turunen, M. *et al.* The effects of UV-B exclusion on the soluble phenolics of young Scots pine seedlings in the Subarctic. *Environ. Pollut.* **106**, 219–228 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Langebartsels, T. Boller, J. Lucht, G. Buchholz, H. Frohnmeyer, P. Crouzet, P. Pelczar, C. Körner and A. Kutenberger for discussions and comments. We also thank S. Stich, V. Gloeckler and C. Ramos for assistance. The photolyase mutant line (*uvr2-1*) was obtained from the *Arabidopsis* Biological Resource Center, Ohio, USA. This work was supported by a fellowship from the Foundation of the Chemical Industry of Basel to G.R. and Novartis Research Foundation.

Correspondence and requests for materials should be addressed to G.R. (e-mail: ries@fmi.ch).

Structure of the dimerized hormone-binding domain of a guanylyl-cyclase-coupled receptor

Focco van den Akker*†, Xiaolun Zhang‡, Masaru Miyagi‡, Xuewen Huo‡, Kunio S. Misono‡ & Vivien C. Yee†‡

*Department of Molecular Biology, ‡Department of Molecular Cardiology and †Structural Biology Center, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, Ohio 44195, USA

The atrial natriuretic peptide (ANP) hormone is secreted by the heart in response to an increase in blood pressure. ANP exhibits several potent anti-hypertensive actions in the kidney, adrenal gland and vascular system. These actions are induced by hormone binding extracellularly to the ANP receptor¹, thereby activating its intracellular guanylyl cyclase domain for the production of cyclic

GMP². Here we present the crystal structure of the glycosylated dimerized hormone-binding domain of the ANP receptor at 2.0-Å resolution. The monomer comprises two interconnected subdomains, each encompassing a central β -sheet flanked by α -helices, and exhibits the type I periplasmic binding protein fold. Dimerization is mediated by the juxtaposition of four parallel helices, arranged two by two, which brings the two protruding carboxy termini into close relative proximity. From affinity labelling and mutagenesis studies, the ANP-binding site maps to the side of the dimer crevice and extends to near the dimer interface. A conserved chloride-binding site is located in the membrane distal domain, and we found that hormone binding is chloride dependent. These studies suggest mechanisms for hormone activation and the allostery of the ANP receptor.

The ANP receptor belongs to the family of membrane guanylyl-cyclase-coupled receptors that include the brain natriuretic peptide receptor, heat-stable enterotoxin receptor and an increasing number of orphan receptors, such as those found in the eye and olfactory system, and is related to the non-coupled clearance receptor^{1,3}. These guanylyl-cyclase-coupled receptors are involved in electrolyte and fluid regulation. The ANP receptor comprises an extracellular hormone-binding domain, a short disulphide hinge region, a transmembrane domain, a kinase-like domain, a coiled-coil domain and the catalytic guanylyl cyclase domain¹. To gain structural insight in the hormonal activation of the ANP and related receptors, we have crystallized its dimerized⁴ glycosylated hormone-binding domain.

The fold of the receptor-binding domain contains two subdomains, each comprising a central predominantly parallel β -sheet flanked by α -helices (Fig. 1a). These subdomains are interconnected with three crossovers located near residues 129, 284 and 374. The binding domain dimer is 125 Å wide and extends 60 Å from the C-terminal residue 435, seven residues from the start of the

transmembrane helix at 442. A 20-Å wide crevice is partially filled with the N-linked glycosylation moieties⁵ attached to N13. The C-terminal region 423–435 forms a protruding irregular structure, including a short disulphide loop.

This ANP receptor hormone-binding domain structure is the first representative of its family and reveals unexpected strong structural similarity to an ancient class of proteins, the bacterial type I periplasmic binding fold protein family. This family includes extracellular solute-binding proteins such as the D-ribose binding and the Leu/Ile/Val binding proteins (LIVBP), and bacterial repressors such as the amidase operon repressor AmiC. These proteins share a common allosteric signalling mechanism: binding of a small molecule such as a sugar, amino acid or ion at the subdomain interface results in a decrease of the inter-subdomain hinge angle. Structurally, the ANP hormone binding domain is most similar to AmiC⁶ and LIVBP⁷ (11% and 18% sequence identity for structurally equivalent residues, respectively), the latter of which shares a conserved disulphide. The r.m.s. deviation for C α atoms drops from 4.7–5.0 Å to 1.7–2.2 Å when superimposing the whole structure and the individual domains, respectively.

In the basal state, the ANP receptor is forced into an inactive dimeric conformation dominated by the intracellular kinase-like domain and coiled-coil region interactions^{2,8,9}. Dimerization of the soluble extracellular domain occurs at high protein concentrations and is enhanced upon hormone binding⁴. The asymmetric unit of the crystal contains a dimerized binding domain (Fig. 1a), in which the C terminus of one monomer is disordered beyond residue 425. The dimer interface contains about 1,800 Å² of total buried solvent-accessible surface and yields a shape complementarity statistic of 0.682; both values are well within the range for physiologically relevant protein–protein interactions^{10,11}. Helices 193–206 and 216–230, and the flanking loop near residue 265 participate in dimer interactions; helices 193–206 of one monomer and 216–230

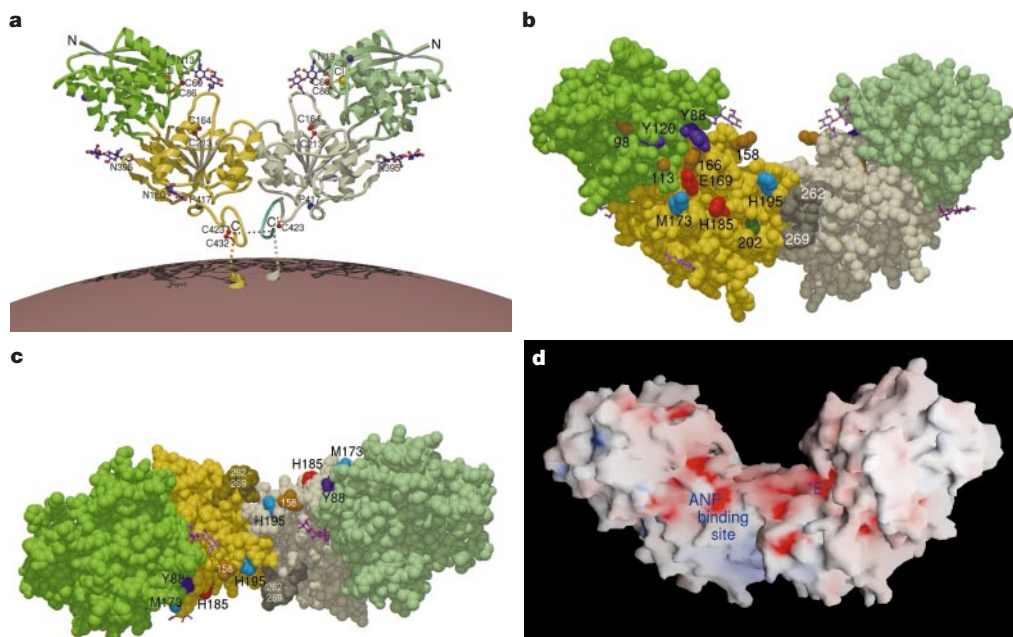


Figure 1 Crystal structure of the ANP receptor hormone-binding domain dimer. **a**, Ribbon diagram showing disulphide bonds (red), glycosylation (purple ball-and-stick), disordered glycosylation sites (purple spheres) and bound chlorides (yellow). The extrapolated C terminus of the second monomer (blue) and two transmembrane helices are modelled. **b**, The ANP-binding site on the receptor, shown as a space-filling model. M173 and H195 (blue), E169 and H185 (red), H185 and A202 (green), and Y88 and Y120 are highlighted. Residues 98, 113, 115, 158, 166 and the common E169 (light brown, 115 is hidden from

view) are structurally equivalent to AmiC residues that interact with AmiR⁶. The dark grey 262–269 loop reaches over from the second monomer and forms a concave surface at one edge of the binding site that may be involved in hormone interactions. **c**, As **b**, but rotated 85° about the horizontal axis. **d**, Electrostatic surface representation generated using GRASP²⁹, with positive and negative charges in blue and red, respectively, and the putative effector pocket labelled 'E'.

of the second monomer run parallel, forming a 2×2 helix bundle. The dimer interface includes eight hydrogen bonds and is equally distributed between hydrophilic and hydrophobic surfaces.

Two prominent features of the crystallized dimer meet the expectations for the extracellular domain in an activated dimer organization presumably allowed by the absence of the suppressive intracellular domains. First, the distance between the two C-terminal Cα atoms (435 in one monomer and 425 in the other monomer) is short, only 25 Å. When residues 426–435 in the first monomer are used to extrapolate the conformation of the C terminus of the second monomer, the distance between the two residues 435 becomes 14 Å (Fig. 1a); these distances are comparable to or shorter than that necessary for transducing signals in other receptors, such as the 21 Å observed for the dimerized erythropoietin receptor bound to its cytokine¹². Experimentally, the two C termini in a constitutively active cysteine mutant ANP receptor were found to be in close proximity¹³. Second, a solvent-exposed, and presumably protease-accessible, conformation is clearly observed for the protruding C-terminal C423–C432 loop, which is stabilized in one monomer by crystal packing and is partially disordered in the second monomer (Fig. 1a). P417 is buried in a hydrophobic pocket and serves as an anchor for the C-terminal region. This local structure is consistent with observations that on ANP binding this loop undergoes a conformational change that renders it protease sensitive, and that P417 is critical for receptor coupling¹⁴.

By affinity labelling studies, we found that ANP residues 4 and 18 are in the vicinity of M173 and the ANP C terminus is near H195 of the receptor (Fig. 1b), in agreement with previous mapping of the hormone-binding site within receptor residues 173–188 and 191–198 (ref. 15). E169K and H185D mutations, both located between M173 and H195 (Fig. 1b), cause loss of hormone binding and confirm this binding site (data to be published elsewhere). The clearance receptor residues 188 and 205, corresponding to H185 and A202 in the ANP receptor, modulate hormonal species specificity¹⁶ and map to the same site (Fig. 1b). The dimeric structure suggests two spatially separated ANP-binding sites, in agreement with the observed 2:2 stoichiometry for the hormone and receptor binding determined by *in vitro* titration experiments⁴ (Fig. 1c). The clefts formed by the dimer are predominantly negatively charged (Fig. 1d), suggesting that electrostatic steering of the basic ANP hormone towards its receptor may be an important determinant of the high binding affinity.

The strong structural similarity of the ANP receptor's binding domain to the bacterial periplasmic binding fold proteins suggests that they have a common allosteric regulation that has not previously been identified in the guanylyl-cyclase-coupled receptors. In AmiC, binding of AmiR is allosterically regulated: acetamide binding causes a subtle hinge motion changing the relative orientation of its two subdomains, thereby releasing the positive regulatory protein AmiR⁶. The hormone-binding site on the ANP receptor corresponds not to the acetamide binding pocket on AmiC, but to its AmiR protein binding site which extends over both subdomains⁶ (Fig. 1b), suggesting that ANP binding could also be allosterically regulated. The inter-subdomain hinge angle in the ANP receptor is more similar to the open conformation of the uncomplexed LIVBP⁷ than to the closed conformation in the AmiC–acetamide complex⁶. Several aspects support a similar hinge based allostery within the ANP receptor. First, a 2° difference in the subdomain orientations is observed between the two monomers of the ANP receptor structure, revealing an inherent hinge flexibility. Second, studies in the ANP¹⁷, clearance¹⁸ and heat-stable enterotoxin¹⁹ receptors highlight critical residues that cause loss of ligand binding when mutated; these residues are all mostly buried and cluster near the inter-subdomain and hinge region. Disturbances in this region may affect the relative subdomain orientation of each monomer, thereby indirectly causing loss of ANP binding. Third, several of the equivalent residues involved in binding acetamide and leucine in AmiC and LIVBP are

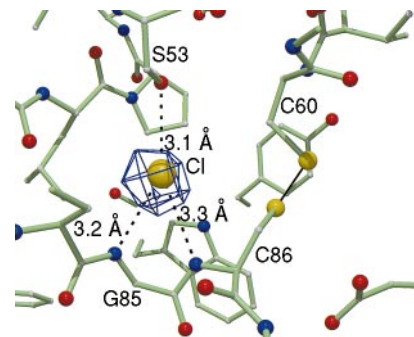


Figure 2 The chloride-binding site in the ligand-binding domain of the ANP receptor. A 25–4-Å resolution anomalous difference Fourier electron density contoured at 5σ is shown in blue.

quite conserved among the hormone-binding receptors, perhaps the consequence of an evolutionary pressure to preserve this allosteric effector binding site, which is a negatively charged pocket in the ANP receptor (Fig. 1d).

Intriguingly, there is a chloride ion buried in the membrane distal domain of each receptor monomer, in the vicinity of the effector site and 8 Å from Y88, which is at or near the hormone-binding site (Figs 1a and 2). The chloride ion is hydrogen-bonded to the backbone N–H moieties of G85 and C86 and to the hydroxyl group of S53, which is strictly conserved in the atrial natriuretic, brain natriuretic, and heat-stable enterotoxin receptors, except for an occasional threonine in other species, suggesting that chloride binding is important. Depletion of chloride from the assay medium results in complete loss of ANP binding to this domain. This chloride dependence can explain enhancement of ANP receptor hormone induced guanylyl cyclase activity in cultured vascular endothelial cells by increases in NaCl concentration²⁰, whereas this activity in the kidney is completely unresponsive to ANP in rats on a very low NaCl diet²¹. Our results suggest a new mechanism of chloride allosteric regulation of hormone binding. For example, the loss of chloride binding under low salt conditions could affect the positions of its liganding atoms, thereby altering the local structure of the receptor near the hormone-binding site rendering it unresponsive to ANP. The structural consequences of chloride depletion on the ANP hormone-binding domain structure not only can

Table 1 Data processing, MIR phasing and model refinement statistics

Crystal	Native	Native	Native	Pt	U
Source	APS	ALS	Raxis4	ALS	Raxis4
Wavelength (Å)	1.000	1.000	1.5418	1.0718	1.5418
Bijvoets merged	Yes	Yes	No	No	No
Unique reflections	80,496	64,638	93,255	81,703	55,455
Redundancy	11.0	7.1	3.4	3.8	2.8
Resolution (Å)	2.00	2.15	2.3	2.45	2.70
<i>R</i> _{sym} (%)	6.6	5.8	5.7	4.6	7.4
	(28.5)	(40.5)	(22.4)	(24.3)	(26.5)
Completeness (%)	98.1	99.5	93.7	98.2	90.0
	(80.9)	(99.6)	(86.2)	(91.2)	(74.3)
Phasing power	—	—	—	1.54	0.44
<i>R</i> _{cullis}	—	—	—	0.66	0.76
Refinement					
Resolution (Å)					100–2.0
Reflections, <i>F</i> > 2σ					78,461
<i>R</i> _{cryst} , <i>R</i> _{free} (%)					19.8, 22.5
r.m.s. deviation					
bond length (Å), angle (°)					0.0104, 1.51
<i>B</i> : bonded main-chain, side-chain atoms (Å ²)					2.41, 3.49
Ramachandran: most favoured, disallowed (% residues)					90, 0

Values in parenthesis are for highest resolution shell. $R_{\text{sym}} = \sum |F_{\text{obs}} - I_{\text{ave}}| / \sum I_{\text{ave}}$, $R_{\text{cullis}} = \sum ||F_{\text{PH}}| - |F_{\text{L}}|| / \sum (|F_{\text{PH}}| + |F_{\text{L}}|)$ for centric reflections, phasing power = r.m.s. $(|F_{\text{PH}}|/E)$, where *E* is the residual lack of closure for acentric reflections. $R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$ and $R_{\text{free}} = R$ factor calculated for test set of reflections not used in refinement (5% comprising 3,965 reflections).

provide a structural basis for the 'emergency brake' mechanism of salt conservation in the kidney in low salt stress conditions, but also may reveal a chloride allosteric mechanism that may be conserved among other guanylyl cyclase coupled receptors.

Note added in proof: Chloride-mediated control of the ANP receptor and its possible role in renal salt regulation and cardiovascular pathophysiology is described in more detail elsewhere³¹. □

Methods

Expression and crystallization

The hormone-binding domain of the ANP receptor was expressed in COS1 cells and purified as described⁴. The protein contains 16% carbohydrate, consisting of N-linked glycosylation at positions 13, 180, 306, 347 and 395 (ref. 5). Protein was concentrated to 10 mg ml⁻¹ in 10 mM HEPES pH 7.4 buffer and crystallized using the hanging-drop vapour diffusion method at 4 °C with drops containing 10 µl each of protein and reservoir consisting of 0.1 M Tris-HCl pH 8.2, 0.2 M Li₂SO₄ and 26% polyethylene 4000 and were stabilized using ethylene glycol before data collection at -180 °C.

Data collection, MIRAS phasing and refinement

Crystals were soaked for 2 days in mother liquor containing 5 mM K₂PtCl₄ at pH 8.4 or in 0.5 mM UO₂(NO₃)₂ at pH 6.9. Optimized anomalous data were collected for the Pt derivative at ALS beamline 5.0.2; data for the U derivative were measured using a Rigaku H3R rotating anode source equipped with Yale mirrors and a RAXIS-4 imaging plate detector. Data were processed using HKL²². The native dataset collected at ALS was used for the MIR phasing calculations, carried out using SOLVE²³. The unit cell has dimensions of 120.3 × 120.3 × 160.6 Å³, with a space group of P4₁2₁2 and 2 molecules forming a dimer in the asymmetric unit. Solvent flattening and twofold averaging using DM²⁴ improved the figure of merit from 0.45 to 0.79, resulting in interpretable electron density for building the structure using O²⁵. Iterative cycles of refinement using CNS²⁶ and the higher resolution native data set collected at APS beamline 19-ID, and model building resulted in an R-factor of 19.8% and R_{free} of 22.5% to 2.0 Å resolution. The presence of chloride ions was verified by collecting a native data set from the same frozen native crystal used at APS with the in-house X-ray source at 1.542 Å, which enhances the anomalous scattering of chloride ions; water molecules have negligible anomalous scattering at this wavelength. Anomalous difference Fourier maps were calculated using data of varying resolution ranges; 7.3 and 5.5. Peaks in the 25–4 Å resolution maps were most prominent and matched the positions of the 2 chloride ions. The temperature factors for the chloride ions are both close to 20 Å² and similar to the temperature factors for its ligands; when the ions were initially modelled as water molecules, unrealistic temperature factors of 2.0 Å² resulted. Interpretable electron density for glycosylation is observed at N13 and N395 in both monomers (GlcNAc-GlcNAc) and N180 (GlcNAc) for one monomer; the remaining sites are disordered. The current model contains protein residues 1–425 of monomer 1, 1–249 and 256–435 of monomer 2, 9 sulphate ions, 2 chloride ions, 9 GlcNAc moieties and 514 waters. Data processing, heavy atom phasing and model refinement statistics are given in Table 1. Figures were generated using MOLSCRIPT²⁷, Raster3D²⁸ and GRASP²⁹.

Characterization of ANP-binding site

The hormone-binding domain was affinity labelled by the stepwise affinity alkylation method³⁰ using ANP peptide analogues containing reactive group at position 4, 18 and 29: N_{4a}-[¹⁴C]iodoacetyl-ANP(4-28); N_{4a}-acetyl-N_{18c}-[¹⁴C]iodoacetyl-[Lys₁₈]ANP(4-28); and N_{4a}-acetyl-N_{29c}-[¹⁴C]iodoacetyl-[Lys₂₉]ANP(4-29). After trypsin digestion, affinity-labelled fragments were isolated by HPLC and sequenced by Edman degradation. The results showed that M173 was crosslinked by the ANP analogues with iodoacetyl group at position 4 and 18, whereas H195 was labelled by the analogue with iodoacetyl group at position 29. Site-directed mutagenesis and ANP-binding studies were carried out in a manner essentially as described¹⁴. Y88 and Y120 are specifically protected by ANP binding from modification by tetranitromethane (data to be published elsewhere). For testing the effect of chloride ion on ANP binding to the hormone-binding domain, binding assays were performed in 20 mM potassium phosphate buffer, pH 7.4, containing 0.1 M concentration of either NaCl, NH₄Cl or CH₃COO NH₄. ANP binding in the presence of NH₄Cl showed 98% specific binding of that in the presence of NaCl, whereas binding decreased to 2% in the presence of CH₃COONH₄.

Received 4 January; accepted 18 April 2000.

- Chinkers, M. *et al.* A membrane form of guanylate cyclase is an atrial natriuretic peptide receptor. *Nature* **338**, 78–83 (1989).
- Foster, D. C., Wedel, B. J., Robinson, S. W. & Garbers, D. L. Mechanisms of regulation and functions of guanylyl cyclases. *Rev. Physiol. Biochem. Pharmacol.* **135**, 1–39 (1999).
- Wedel, B. J. & Garbers, D. L. New insights on the functions of the guanylyl cyclase receptors. *FEBS Lett.* **410**, 29–33 (1997).
- Misono, K. S., Sivasubramanian, N., Berkner, K. & Zhang, X. Expression and purification of the extracellular ligand-binding domain of the atrial natriuretic peptide (ANP) receptor: monovalent binding with ANP induces 2:2 complexes. *Biochemistry* **38**, 516–523 (1999).
- Miyagi, M. & Misono, K. S. Identification and characterization of glycosylation sites in the

extracellular domain of the atrial natriuretic peptide receptor: Oligosaccharide structures are not required for hormone binding. *Eur. J. Biochem.* (submitted).

- O'Hara, B. P. *et al.* Crystal structure and induction mechanism of AmiC-AmiR: a ligand-regulated transcription antitermination complex. *EMBO J.* **18**, 5175–5186 (1999).
- Saper, M. A. & Quirocho, F. A. Leucine, isoleucine, valine-binding protein from *Escherichia coli*. Structure at 3.0-Å resolution and location of the binding site. *J. Biol. Chem.* **258**, 11057–11062 (1983).
- Wilson, E. M. & Chinkers, M. Identification of sequences mediating guanylyl cyclase dimerization. *Biochemistry* **34**, 4696–4701 (1995).
- Lowe, D. G. Human natriuretic peptide receptor-A guanylyl cyclase is self-associated prior to hormone binding. *Biochemistry* **31**, 10421–10425 (1992).
- Conte, L. L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Syed, R. S. *et al.* Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature* **395**, 511–516 (1998).
- Labrecque, J., McNicoll, N., Marquis, M., & De Léan, A. A disulphide-bridged mutant of natriuretic peptide receptor-A displays constitutive activity. *J. Biol. Chem.* **274**, 9752–9759 (1999).
- Huo, X., Abe, T. & Misono, K. S. Ligand binding-dependent limited proteolysis of the atrial natriuretic peptide receptor: juxtamembrane hinge structure essential for transmembrane signal transduction. *Biochemistry* **38**, 16941–16951 (1999).
- McNicoll, N., Gagnon, J., Rondeau, J.-J., Ong, H. & De Léan, A. Localization by photoaffinity labeling of natriuretic peptide receptor-A binding domain. *Biochemistry* **35**, 12950–12956 (1996).
- Engel, A. M. & Lowe, D. G. Characterization of the hormone binding site of natriuretic peptide receptor-C. *FEBS Lett.* **360**, 169–172 (1995).
- Duda, T., Goraczniak, R. M. & Sharma, R. K. Site-directed mutational analysis of a membrane guanylate cyclase cDNA reveals the atrial natriuretic factor signaling site. *Proc. Natl Acad. Sci. USA* **88**, 7882–7886 (1991).
- Itakura, M., Suzuki, H. & Hirose, S. Structural analysis of natriuretic peptide receptor-C by truncation and site-directed mutagenesis. *Biochem. J.* **322**, 585–590 (1997).
- Wada, A. *et al.* Identification of ligand recognition sites in heat-stable enterotoxin receptor, membrane-associated guanylyl cyclase C by site-directed mutational analysis. *Inf. Immun.* **64**, 5144–5150 (1996).
- Katafuchi, T. *et al.* Modulation by NaCl of atrial natriuretic peptide receptor levels and cyclic GMP responsiveness to atrial natriuretic peptide of cultured vascular endothelial cells. *J. Biol. Chem.* **267**, 7624–7629 (1992).
- Veress, A. T., Honrath, U., Chong, C. K. & Sonnenberg, H. Renal resistance to ANF in salt-depleted rats is independent of sympathetic or ANG-aldosterone systems. *Am. J. Physiol.* **272**, F545–F550 (1997).
- Otinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- The CCP4 suite: Programs for computational crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Kraulis, P. J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystal.* **24**, 946–950 (1991).
- Merritt, E. M. & Bacon, D. J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
- Nicholls, A., Sharp, K. A., & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
- He, X., Nishio, K., & Misono, K. S. High-yield affinity alkylation of the atrial natriuretic factor receptor binding site. *Bioconjug. Chem.* **6**, 541–548 (1995).
- Misono, K. S. Atrial natriuretic factor binding to its receptor is dependent on chloride concentration: a possible feedback-control mechanism in renal salt regulation. *Circ. Res.* **86**, 1135–1139 (2000).

Acknowledgements

This work was supported by grants from the NIH (K.S.M.) and the American Heart Association (K.S.M. and V.C.Y.) and by a NIH postdoctoral fellowship (F.v.d.A.). We thank G. Stark for his support. Diffraction data were measured at the Advanced Light Source and on the Structural Biology Center beamline at the Advanced Photon Source, supported by the US Department of Energy, and at the National Synchrotron Light Source, Brookhaven National Laboratory, supported by the U.S. Department of Energy and the National Institutes of Health. We are grateful to E. Walker and the CCF LRI Computer Core for facilities support, to S. Ginell and T. Earnest for beamline support, to L. Pearl for coordinates of the AmiC/AmiR structure, and to M. Young for interesting discussions. Purification and crystallization of the ANP receptor hormone-binding domain, ANP-binding site determination, binding studies and other functional characterizations were carried out by X.Z., M.M., X.H. and K.S.M. Data collection, structure determination and refinement, and structure analysis were carried out by F.v.d.A. and V.C.Y.

Correspondence and requests for materials should be addressed to V.C.Y. (e-mail: yeev@ccf.org) or K.S.M. (e-mail: misonok@ccf.org). Coordinates and diffraction data have been deposited with the RCSB PDB (accession number 1DP4).



Great expectations

New buildings are springing up to house cross-disciplinary research
p107



The 'in' thing

Benchwork is now being married with traditional engineering
p108

Laying a firm foundation for interdisciplinary research endeavours

US universities and institutes are becoming increasingly aware of the benefits of a multidisciplinary approach to research, says Diane Gershon.

Interdisciplinary research is not a new concept, but such initiatives are increasing in number in the United States. For example, the importance of research that transcends the traditional research boundaries had been recognized for some time at the University of Pennsylvania (Penn). But the university put the foundations for its interdisciplinary biomedical research infrastructure in place without much of a fanfare, explains Peter Davies, director of the Institute for Medicine and Engineering (IME).

In 1996, the university formed the IME to link the schools of medicine and engineering in what was "a top-down decision", says Davies. The university has since established a Center for Bioinformatics and a Center for Bioactive Materials — both operating under the umbrella of the IME — and a new Center for Medical Informatics is planned.

A lot of biomedical research is still highly descriptive, says Davies, but the trend is now to use quantitative and computational approaches to solving biological problems. He sees the institute's primary role as providing a flexible and dynamic interface between the worlds of medicine and engineering. No easy task given the differences between the disciplines — one is largely hypothesis-driven, the other driven by design.

Medicine meets engineering

In 1997, the IME moved into a new building next to the university hospital and the engineering and medical schools. Over the past three years, 11 faculty have been recruited — about half through the engineering school and half through the medical school — most of whom, says Davies, have dual training in both the biological and the more quantitative sciences. Dennis Discher, for example, who studies membrane and cell mechanics, did his first degree in engineering, his PhD in bioengineering and his postdoctoral work in biophysics. Davies says Discher can flip back and forth between the more quantitative biophysical/engineering aspects of his work to studies involving a more molecular, protein chemistry approach.



Building for the future: the new bioengineering and bioscience building at Georgia Tech (above) and an aerial view of the interdisciplinary complex.

In addition to the 11 core faculty, the institute has a further 52 members drawn from the broader Penn community — all of whom are involved in cross-disciplinary work on some level.

Davies, who was lured from the University of Chicago in late 1995 to head the IME, says he has deliberately kept the definition of the kind of research that should go on in the institute fairly broad. But he does stipulate that the work has to be fundamental and have a strong quantitative component. With its current composition, the IME has considerable strength in interdisciplinary cardiovascular research.

Despite a temporary freeze on academic development funds because of last year's \$200 million budget deficit for Penn's academic medical centre, Davies says that the IME has managed to recruit several interdisciplinary faculty recently using resources fronted principally by the departments. Biophysicist Paul Janmey was recruited from Harvard to lead a cross-institute effort in cytoskeletal and biomaterials research, and Valerie Weaver has come from the Lawrence Berkeley National Laboratory to work on cancer metastasis.



Looking forward, a major challenge is to get big donors to provide substantial support to expand the IME's facilities, says Davies. As a first step, the IME has joined with the Department of Bioengineering to submit a proposal to the Whitaker Foundation for support to build a new home for bioengineering, which would include some space for the IME.

How to retain key people given the current proliferation of cross-disciplinary research initiatives is another area of

careers and recruitment

concern, says Davies. Whereas the university moved quickly and aggressively to secure some of the best people in 1996 and 1997, these sorts of individuals are in high demand and are targets for recruitment elsewhere. To some extent, the university has been hiding its light under a bushel, says Davies, who has been given the task of pulling together under one umbrella — albeit a virtual one — information on all of the cross-disciplinary research activities at Penn. This soon-to-be-launched website will be called the Biological Engineering Network at Penn.

Georgia takes the initiative

In a similar move to Penn, the Georgia Institute of Technology has embarked on an ambitious infrastructure development programme. “Many universities pay lip-service to interdisciplinary endeavours,” the institute’s president Wayne Clough told faculty and staff last October, “but Georgia Tech is actually becoming interdisciplinary from the ground up.” The initiative involves building three interdisciplinary research ‘neighbourhoods’, or complexes, designed to encourage exchanges and interactions between faculty members and students from scientific and engineering disciplines.

The building programme is a major expansion for the institute. The last science-related facility to be built on campus was constructed some 30 years ago, when Georgia Tech was primarily an undergraduate teaching establishment for engineering students. Since then the institute has gone from having virtually no dollars for research to having over \$220 million for this year.

The new complex is known as BEM: Biotechnology–Environmental Science and Technology–Molecular and Materials Science. The first building to go up was the \$32 million Institute for Bioengineering and Bioscience, which opened its doors last summer and brings together biochemists, bioengineers and biologists. Michael Thomas, provost and vice-president of academic affairs at Georgia Tech, says work will soon start on the \$54 million Environmental Science and Technology building. This will house most of the chemical engineering faculty, the school of Earth and atmospheric sciences, civil engineers concerned with low-level ozone generation, as well as environmental biologists and analytical chemists.

A fund-raising effort is underway for a third building in the complex, for the School of Biomedical Engineering and the joint Georgia Tech/Emory University academic and research programme in biomedical engineering. Georgia Tech and Emory University School of Medicine linked up in 1997 to form a joint department of biomedical engineering. This unusual partnership between a public and a private institution situated five miles apart is mutually beneficial as Emory does not have an engineering school



Spirit of cooperation: Penn pushes a multidisciplinary approach to research.

and Georgia Tech lacks a medical school. Despite having to work through issues relating to promotion and tenure beforehand, formation of the joint department “gave us a way to take what’s complementary in both places, put it together and make something bigger”, says Thomas. The department currently has nine faculty but this is set to grow to 22 within four years, Thomas says.

Construction of a fourth building, dedicated to molecular and materials science and engineering, will begin in about two years. A second research complex devoted to manufacturing research has just been completed, and there are plans for a third in the area of information technology, which will bring together faculty from the College of Computing and the School of Electrical and Computer Engineering.

Bridging the divide in Chicago

Four years ago, Glenn Steele Jr, dean of the division of biological sciences and vice-president for medical affairs at the University of Chicago, and David Oxtoby, dean of the division of physical sciences, resuscitated the idea for an interdisciplinary research initiative on campus. This renewed interest in promoting activities at the interface of the physical and biological sciences led to the formation of the Institute for Biophysical Dynamics. At full strength, this institute will have 24 core faculty members (12 from each division) and will bring computer scientists, theoreticians and experimentalists in the biological sciences, chemistry and physics together under one roof.

Steele says the institute was set up with the proviso that most of its faculty members would be new recruits. “We wanted to use this as a way of recruiting people that we couldn’t have recruited otherwise,” he says. About one-third of the slots are already filled, even though the foundations for the institute’s new home have yet to be laid.

The new institute will eventually reside in the Interdivisional Research Building, which should be finished by 2003 at a cost of between \$170 million and \$180 million. In addition to the Institute for Biophysical Dynamics, the building will house scientists from the Department of Biochemistry and Molecular Biology, the Ben May Institute for Cancer Research and the Department of Chemistry, as well as physicists from the James Franck Institute. This January, the Howard Hughes Medical Institute (HHMI) agreed to contribute \$17.6 million to the construction costs for the building, and one floor of it will now be occupied by the nine HHMI-supported investigators currently at the university. The HHMI’s involvement and willingness to be part of the new building was the “icing on the cake”, says Steele.

Much about the new Institute for Biophysical Dynamics is symmetrical. There are two co-directors: Norbert Scherer, recruited from the University of Pennsylvania, who heads the physical sciences, and Anthony Kossiakoff, formerly of Genentech, who leads the biological sciences. Moreover, each division is footing half the bill.

The intention is to use the new institute to promote and catalyse cross-disciplinary exchanges between the many researchers who will eventually reside in the new building, as well as with researchers in the larger university community, the Argonne National Laboratory and in industry. When it is built, the new Interdivisional Research Building will be at the nexus of two science quads, which are surrounded by facilities for the physical and biological sciences.

The changing climate in support of a multidisciplinary approach to research has meant that research proposals at the interface between disciplines, which five or ten years ago would have been deemed too risky, are now being funded. Institutions across the United States are taking advantage of this climate, and are examining how best to position themselves for the future in a way that maximizes their key strengths and creates an enhanced environment for interdisciplinary research and training. They are building new facilities that promote as much interaction as possible and are finding these buildings and the programmes themselves a considerable inducement for attracting the brightest and the best to their research campuses. ■

Diane Gershon is Assistant Editor, New Technology, Nature Medicine.

University of Pennsylvania Institute for Medicine and Engineering

♦ <http://www.med.upenn.edu/ime>

University of Pennsylvania Center for Bioinformatics

♦ <http://www.pcbi.upenn.edu/>

Georgia Institute of Technology

♦ <http://www.gatech.edu/>

Georgia Tech/Emory joint Department of Biomedical Engineering

♦ <http://www.bme.gatech.edu/>

University of Chicago, Institute for Biophysical Dynamics

♦ <http://bmb.bsd.uchicago.edu/IBDHome.html>